Phase transitions in Bak-Sneppen avalanches and in a continuum percolation model

Gillett, A.J.

2007

Phase transitions in Bak-Sneppen avalanches and in a continuum percolation model

Alexis Jonathan Gillett

ISBN: 9789086591305
The circular cover design is inspired by the dynamic lilypond model. The main image on the front cover is a realisation of a Bak-Sneppen model. Each spoke represents one species with the length of the spoke denoting its fitness. The other images inside the circles are realisations of dependent continuum percolation models.
On the outside bottom corner of each page is an image. Together these images form a flick book. When turning the pages rapidly, these pictures create an animation. Starting from the front of the book, there is a realisation of a Bak-Sneppen model. Starting from the back of the book, there is a realisation of the generalised nearest neighbour model with varying α.

Thomas Stieltjes Institute
for Mathematics

VRIJE UNIVERSITEIT

Phase transitions in Bak-Sneppen avalanches and in a continuum percolation model

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Exacte Wetenschappen
op maandag 17 september 2007 om 13.45 uur
in het auditorium van de universiteit,
De Boelelaan 1105

door

Alexis Jonathan Gillett

geboren te Harrogate, Engeland

promotor: prof.dr. R.W.J. Meester

Contents

List of Figures

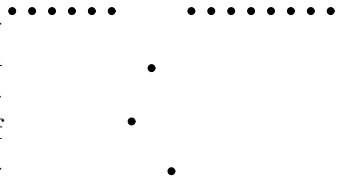
Chapter 1

Introduction

This thesis concerns itself with the rigorous analysis of two stochastic models. Since both models will be defined formally in the later chapters of this thesis, this introductory chapter will give background information about these models that might be of interest to and understandable by non-mathematicians. The rest of this thesis takes the following form. Chapter 2 establishes some fundamental properties of the Bak-Sneppen model and acts as a summary of known results about the model. Chapters 3 and 4 go into more details for some of these results. Chapter 5 is about a new continuum percolation model and is based on the article that first introduced this model.

1.1 Stochastic Models

A stochastic or probabilistic model is a mathematical model containing an element of randomness. A simple example of this is a model for the outcome of tossing a coin. The usual way to model this is to say that the outcome is a head with a 50% chance (i.e. with probability $\frac{1}{2}$), otherwise it is a tail (also with probability $\frac{1}{2}$). Intuitively this seems like a very reasonable model and there has even been some research confirming its validity (in usual circumstances) [28]. Why is such a model so natural? Surely the behaviour of a coin toss is determined by the physical laws of motion and is, therefore, not 'random'. Yet a stochastic model still seems highly appropriate. In the absence of sufficient information, such as the initial conditions of the coin toss, basing the model on the statistical properties of a coin toss is the most sensible option. This simple example demonstrates the wide-ranging relevance of stochastic models. In situations with only imprecise or incomplete data, a stochastic model seems



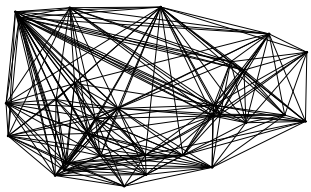
more reasonable than a deterministic one. Importantly, no belief about some fundamental 'randomness' of the phenomena under study is required.

My specific area of research is interacting particle systems, which is also sometimes referred to as spatial stochastics or statistical mechanics. This involves the study of stochastic models where there is a number of particles that interact with each other. This interaction is normally conveniently described via some sort of spatial structure. Many of the models studied by mathematicians working in this field are originally motivated by the work of theoretical physicists, where a physical system is modelled by defining interactions between microscopic particles. The main interest is then on the macroscopic / global properties of the system. For example, one might consider a model of a ferromagnet. Here the model could be based on the charge of individual atoms and the interactions between them. However, the main quantity of interest is the overall magnetism of the ferromagnet.

The common feature of this area of study is the existence of two different scales: the microscopic and the macroscopic. A model is defined in terms of the behaviour of the microscopic particles and the goal to calculate the behaviour of the system as a whole, i.e. at the global or macroscopic level. The microscopic particles can represent anything from atoms to people. A typical feature is that the macroscopic world is made up from an extremely large number of the microscopic particles. So large in fact, that one treats the macroscopic world as consisting of infinitely many microscopic particles.

1.2 Phase Transitions

Recall that we are interested in studying the overall behaviour of models that are defined in terms of the stochastic interactions between a large number of small particles, which together make up the system. Since the model itself is defined by some stochastic rules, it might seem logical to expect the behaviour of the model as a whole to be stochastic. However, this is often not the case. Situations where randomness at the microscopic level does not transfer to the macroscopic level are very interesting to mathematicians. This property might seem like a paradox at first, but the existence of non-trivial determinism at the macroscopic level arises because it is made up of infinitely many microscopic particles.



Often the behaviour at the microscopic level is determined by some parameter and it is then interesting to see how different parameter values affect the behaviour of the model. In particular, one is interested in seeing if there is a phase transition. A phase transition occurs when varying the

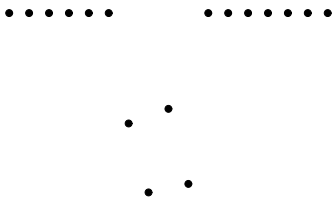
parameter leads to a sudden change in the qualitative behaviour of the system as a whole. A good example of a phase transition is how ice changes to water with the parameter in this case being the temperature. A model exhibiting a phase transition is very interesting because small changes in the parameter value can have dramatic effects changing the nature of the deterministic behaviour at the macroscopic level. Typically models have only one phase transition, which divides the range of parameter values into two phases: the subcritical phase and the supercritical phase. The model discussed in Chapter 5 of this thesis is a good example of a spatial stochastic model exhibiting a phase transition.

1.3 A near-neighbour continuum percolation model

Percolation is one of the classical spatial stochastic models. It can be used to study the flow of fluid through a semi-porous substance or to model networks. In standard percolation, particles are arranged on a grid. Each particle is connected to its neighbouring particle with some probability. For example, the connections could be independent of each other and with probability p. The overall connectivity of the model depends on the choice of p. At the global level this is examined by asking whether there is an infinite collection of particles that are connected to each other (this is known as an infinite cluster). As p is varied, the model exhibits a phase transition. For a fixed choice of p, there is either always an infinite cluster or never an infinite cluster.

The model considered in Chapter 5 is similar to the model described above. However, there are two important differences. Firstly, the particles are not confined to a grid, but occupy a continuous space. Secondly, the connections between the particles are not independent of each other. These dependencies make the model more difficult to analyse, but such dependencies often make the model more applicable to the modelling of real life situations. The basic question, though, is still the same. For which parameter values is there an infinite cluster?

This model does exhibit a phase transition, i.e. for some values there is always an infinite cluster and for some values there is never an infinite cluster. This and related results about the phase transition can be found in Chapter 5. Unfortunately it is not possible to show an infinite cluster in a (finite) picture. However, Figure 1.1 does give you a flavour of what the phase transition looks like in this case. In the subcritical regime (the picture on the left), there are lots of small isolated



networks. In the supercritical regime (the picture on the right), nearly all the particles are contained in one large global network.

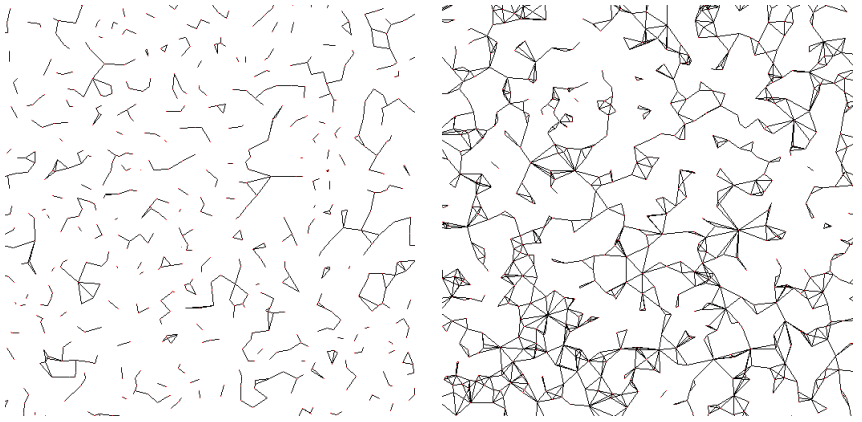
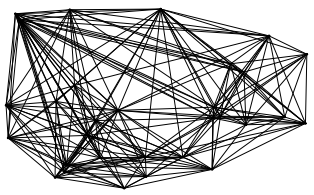


Figure 1.1: The generalised nearest neighbour model

1.4 Criticality

A model is said to be critical when it has its parameter set to be on the boundary between two different phases. A critical model has a very interesting property called scale invariance. This can be expressed in terms of equations (power laws) and shapes (fractals). A fractal is an object that looks roughly the same under different levels of magnification, i.e. something that is more or less made up of miniature copies of itself. As long as you allow for a minimum scale, it is easy to find examples of fractals in nature. Figure 1.2 is a magnified image of a vegetable, showing its fractal structure. Other commonly sited examples of fractals are coastlines with the coastline paradox stating the length of a coastline depends on how you measure it. Due to the fractal nature of coastlines, the length of a coastline increases as you attempt to measure it more accurately.



A power law, $p(x) \propto x^{-\tau}$, is a scale-invariant distribution. In words, this formula states that the probability of observing a given value varies inversely as a power of that value. Replacing x with some multiple of x, i.e. rescaling the model, does not alter the shape of the distribution. Many physical phenomena appear to obey power law distributions. Perhaps the most famous example of

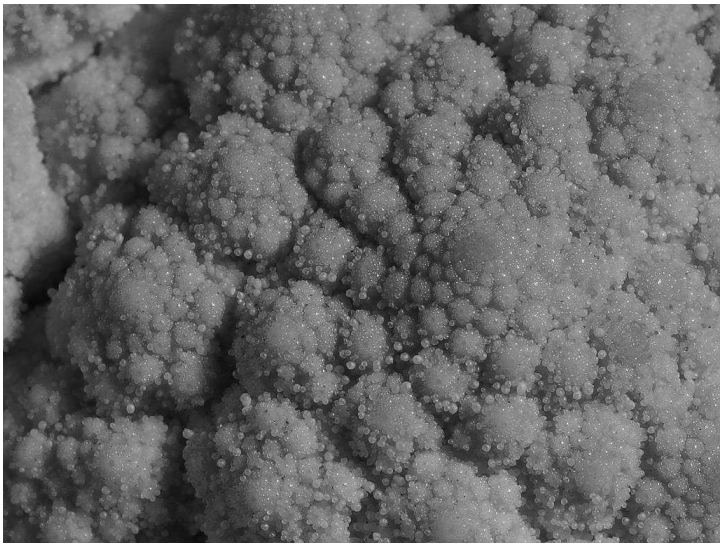


Figure 1.2: The fractal structure of a broccoflower

a power law distribution is the Gutenberg-Richter law, which relates the frequency of earthquakes to their magnitude. A power law can be elegantly represented as a straight line on a log-log plot, see Figure 1.3.

In classical models, one only achieves scale-invariance at criticality. The parameter of the model must be tuned precisely at the point of a phase transition, otherwise fractals and power laws aren't observed. However, scale-invariance is fairly ubiquitous in nature; many natural phenomena behave like critical systems. This gives rise to a fundamental problem. If these classical models can truly explain these real world phenomena, why are they so commonly tuned to their critical values?

1.5 Self-organised Criticality

Self-organised criticality has been proposed as an explanation for the prevalence of scale-invariance. The idea dates back to a 1987 letter by by Bak, Tang and Wiesenfeld [3]. The authors had been inspired by earlier work [2] examining a mechanical system consisting of a large number of linked pendulums. This system was attracted to critical states, namely where slightly moving one pendulum could have a knock-on effect changing the entire system. This was, at the time, rather surpris-

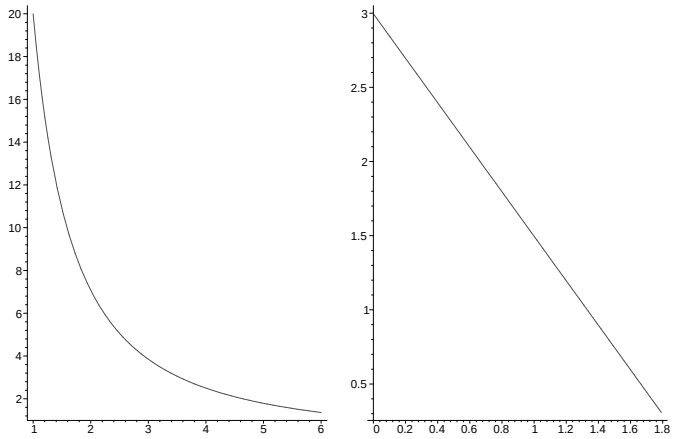
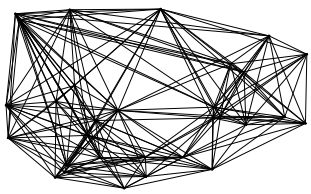


Figure 1.3: A power law displayed on a normal plot and a log-log plot

ing. To see why, consider just one pendulum. Due to gravity, the pendulum can theoretically only stop in two positions: vertically up or vertically down. The behaviour of the linked pendulums was rather like the single pendulum having a tendency to end up in the vertically up position!

In their letter 'Self-Organized Criticality: An Explanation of $1/f$ Noise' [3] and subsequent paper 'Self-Organised Criticality' [4], Bak, Tang and Wiesenfeld proposed an explanation for the complex behaviour observed in many physical systems. The idea was that these systems were, like the linked pendulums, attracted to their critical states where small changes to the system could have global consequences. In other words, they organised themselves to their critical states. Going back to the example of percolation models, the suggestion was that these systems self-tuned their parameters over time to their critical values.

A more elegant example of a self-organised critical (SOC) model than the linked pendulum model was soon devised: the Abelian sandpile model. This is probably the most famous example of self-organised criticality. The Bak-Sneppen model is another well-known example. In general, SOC models are defined as having no tuning parameter, although it is usually possible to recover some sort of hidden parameter that is self-tuned by the model. Self-organised criticality has attracted a lot of attention in the scientific literature. Per Bak's own book on the subject is boldly entitled *How Nature Works* [1].



The essential feature of self-organised criticality is spontaneous critical behaviour such as power

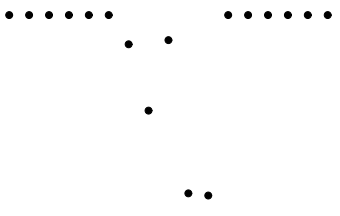
laws and fractals. This is in stark contrast to classical criticality where the critical behaviour is less robust, requiring precise tuning of parameters. There has been some philosophical discussion about the fundamental differences between SOC and classical models. A recipe for cooking up SOC models from classical models has been proposed [40]. The same paper claimed that for SOC models the specification of the rules of the model amounted to some sort of tuning of the model. In their conclusion, the author's hinted at the perhaps important idea that SOC models should really be considered as a classical system with an external driving force that leads it to criticality. Rigorous connections have been drawn between classical and SOC systems [41], showing that the rules of an SOC system naturally force it to a classical system at criticality.

1.6 The Bak-Sneppen model

The Bak-Sneppen model was originally introduced as a simple model of evolution by Per Bak and Kim Sneppen in 1993 [5]. Their model can be defined as follows. There are N species arranged on a network, each allocated a value called a fitness. A species' fitness represents how well-adapted it is to its environment. At each time step, the lowest fitness is located and replaced by a new random fitness. The fitnesses of any neighbouring species are also replaced. The justification for replacing the lowest fitness is a sort of *survival of the fittest* paradigm. The replacement of the neighbouring fitness is supposed to represent the interactions between different species.

This is a very simplified model of evolution with the essence of each species condensed into just one value, its fitness. However, the goal was not to create a very detailed model that tried to capture as many features of evolution as possible. Instead, the aim was to devise a very simple abstraction of an ecosystem, which still captured some features of evolution and that was self-organised critical. Various sources claim that extinction events seem to obey a power law distribution, see for example [9] and references. The existence of power laws suggested that evolution could be an example of self-organised criticality.

From a mathematical point of view, the model is very interesting. The rules of the model are simple enough to permit rigorous analysis, but the model still displays complex behaviour. Chapters 2, 3 and 4 give a mathematical treatment of this model.



1.7 Outline of the Thesis

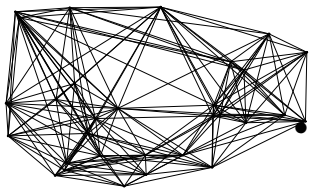
The rest of this thesis consists of four separate articles, reproduced here more or less in their original form. Therefore, it is possible to read the remaining chapters in any order. However, Chapter 2 gives a mathematical introduction to the Bak-Sneppen model. Therefore it would be logical to read this chapter before attempting Chapters 3 and 4, which are also about the Bak-Sneppen model. Since Chapter 2 serves as a sort of summary of the results known about the Bak-Sneppen model, some of the results listed here were proven by other authors. In particular, a number of these results in this chapter are due to my predecessor at the Vrije Univesiteit, Dimitri Znamenski working in collaboration with Ronald Meester [37, 38].

In Chapter 3, the maximal avalanche decomposition of the Bak-Sneppen model is examined. With SOC models, attention is normally directed towards assessing the nature of the critical behaviour. However, this chapter looks how the model evolves towards criticality. We look at this by considering the individual avalanches of the maximal avalanche decomposition. Chapter 4 derives bounds for the critical values of Bak-Sneppen avalanches on infinite graphs. In particular, the critical values are shown to be non-trivial on a wide range of graphs.

Chapter 5 introduces a new continuum percolation model. The model introduced is a generalisation of an existing dependent percolation model: the nearest-neighbour model. The main focus of this chapter is on the critical values of the model. The thesis then concludes with a couple of short appendices, a bibliography, acknowledgements and a summary of the thesis in Dutch.

1.8 List of Publications

- **Maximal avalanches in the Bak-Sneppen model** with Ronald Meester and Peter van der Wal. *Journal of Applied Probability*, volume 43, number 3, pages 840-851, 2006



- **Bounds for avalanche critical values of the Bak-Sneppen model** with Ronald Meester and Misja Nuyens. *Markov Processes and Related Fields*, volume 12, number 4, pages 679-694, 2006
- **A near-neighbour continuum percolation model** with Misja Nuyens. *Submitted*

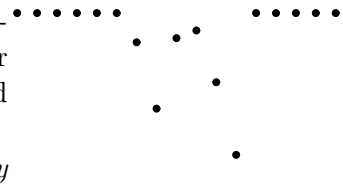
Chapter 2

A mathematical introduction to the Bak-Sneppen model

This chapter formally defines the Bak-Sneppen model and contains a summary of the mathematical results about the model known to date. Alongside some new results, in Section 2.4 particularly, basic techniques are explained and used to establish some fundamental properties of the model.

2.1 Introduction

The Bak-Sneppen (BS) model was introduced by two physicists, Per Bak and Kim Sneppen, in 1993 [5]. It had been observed that certain phenomena relating to evolution appeared to be scale free with behaviour governed by power law distributions [9]. Power laws typically arise in mathematical models at criticality, which normally only occurs when the parameters of the model are set to specific values (tuning). For example, power laws only occur in ordinary percolation when the open bond probability is exactly equal to its critical value. Bak, together with Tang and Wiesenfeld, had introduced, in 1987, the idea of self-organised criticality (SOC) [3, 4]. SOC models behave like models at criticality without the need to fine tune parameters, in a sense they are self-tuning. The Bak-Sneppen model was devised as a simple SOC model of evolution. Although evolution is considered by many as an example of SOC, it has been argued that the power law behaviour of evolution can also be explained by non-SOC models [12].



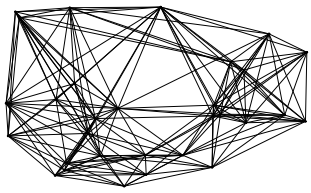
The Bak-Sneppen model is essentially a *toy* model of evolution. Its rules were chosen not so

much to accurately model real evolution, but more to give a flavour of the dynamics that underpin evolution whilst keeping the model simple enough to analyse. Despite this the model has been applied to bacterial evolution [22, 23]. The BS model does not have to be exclusively thought of as an evolution model and works equally well as a model of any sort of phenomena where a survival of the fittest (actually extinction of the weakest) paradigm is in place. For example, the BS model has also been applied in economics to model the competition of companies in a market [20, 21].

The Bak-Sneppen model has received a fair amount of interest in the scientific literature. The vast majority of this attention has come from the theoretical physics community. A number of results about the model are now available, see for example [10], but these would be considered non-rigorous by mathematicians. Results are typically obtained via numerical simulations or heuristic arguments (sometimes referred to as analytic methods). Many variants of the model have also been devised, since modifying the rules of the model is a simple task if you are simulating the model on a computer (see Section 2.1.4).

Ignoring the physical motivation for the model, the Bak-Sneppen model is interesting from a purely mathematical perspective since it has similarities with well-known models such as site percolation, branching processes, the contact process and random walks. Like many such models, the rules of the Bak-Sneppen model are simple to define, but lead to highly non-trivial behaviour. However, rather little has been achieved from a rigorous mathematical perspective. To date only four mathematical articles have been published about the BS model [37, 38, 42, 43]. A similar list of physics articles about the model would be prohibitively long. Instead, readers are directed to the references for a non-exhaustive list.

One of the aims of this chapter is to encourage more research on this model by providing a solid mathematical basis and framework for the Bak-Sneppen model. When two separate fields work on the same problem, collaboration between the two is sometimes rather limited. Thus a further aim is to introduce mathematicians to the relevant physical literature about the BS model. There are many interesting conjectures and open problems for the Bak-Sneppen model. This chapter takes the following form. After defining the model, the fundamental problems of the model will be introduced accompanied by a motivating example and further context. The main technique for analysing the model will be introduced followed by more advanced results. This will take the form of a mixture of a summary of previous results and new results. We shall con-



clude with a re-cap of the key open problems for the model and with suggestions for further research.

2.1.1 Definition

We begin with a formal definition of the Bak-Sneppen model. This is different to the original model presented in [5], but is a natural generalisation that includes the original model as a special case. Definition 2.1 defines the state space of the Bak-Sneppen model.

Definition 2.1. *We call a vector decent if it has a unique minimal component. For G a graph with vertex set $V(G)$, $\Omega_G = \{\omega \in [0,1]^{V(G)} : \omega \text{ is decent}\}$.*

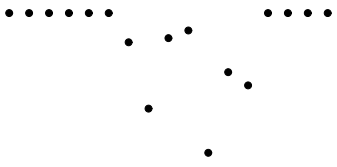
Since the dynamics of the Bak-Sneppen model are governed by the location of the minimal component of some vector, the model is only well-defined if we restrict ourselves to Ω_G.

Definition 2.2. *The Bak-Sneppen model on a locally finite connected graph, G, with initial fitness distribution π is a discrete time stochastic process, $\{\Phi_n\}_{n\geq 0}$ on the state space Ω_G. The initial state is determined by a probability distribution on Ω_G, π , i.e. $\Phi_0 \sim \pi$. For $v \in V(G)$, the value $\Phi_n(v)$ is referred to as the fitness of v at time n. Let v_n^M denote the (random) location of the minimal component of Φ_n and $\Gamma^\star(v) = \Gamma(v) \cup v$, where $\Gamma(v)$ denotes the neighbours of the vertex v in the graph G. Then*

$$\Phi_{n+1}(v) = \begin{cases} U_{n+1}^v & \text{if } v \in \Gamma^\star(v_n^M) \\ \Phi_n(v) & \text{otherwise,} \end{cases}$$

where U_{n+1}^v is a uniform $(0,1)$ random variable independent of everything else. If $\Phi_{n+1} \notin \Omega_G$, then the process terminates.

The minimal value $\min_{v\in V(G)} \Phi_n(v)$ is referred to as the *minimal fitness*, with its location, v_n^M, being called the *minimal vertex*. It is typical to distinguish one vertex of the graph by calling it the origin. Definition 2.2 restricts us to locally finite connected graphs. There is little value in considering disconnected graphs, since such a model would behave like a separate Bak-Sneppen model on each connected component of the graph. Henceforth, any graph mentioned should be assumed to be connected, unless explicitly stated otherwise. The need to consider only locally finite graphs (i.e. graphs where every vertex has finite degree)

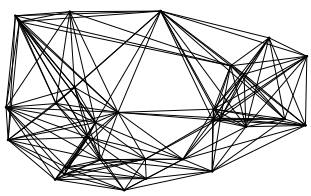


has to do with the well-definition of the model and this topic is dealt with in the next paragraph. The original model of Bak and Sneppen can be recovered by setting $G = \Lambda_N$, where Λ_N denotes the circular graph of order N) and π to the N-dimensional product measure of the uniform $(0, 1)$ distribution.

Since each update step of the Bak-Sneppen model requires finding the minimal fitness, the model is only well-defined on decent configurations. Note that $\Phi_n \in \Omega_G \not\Rightarrow \Phi_{n+1} \in \Omega_G$. Ideally you would like to be able to guarantee that the model never terminates, i.e. it always remains in Ω_G. There are two things that can go wrong. Firstly the minimal fitness might not exist and secondly the minimal fitness might not be unique. Since any update of a vertex of infinite degree leads the model to have no minimal fitness, it is natural for the definition of the model to exclude such possibilities. On a locally finite graph, the update procedure of the model almost surely creates no tied fitnesses and the minimum fitness from amongst the updated vertices always exists. Therefore, any problems with the model stem from the initial configuration.

We call an initial fitness configuration, $\omega \in \Omega_G$, *reasonable* if $\Phi_0 = \omega \Rightarrow \Phi_n \in \Omega_G \; \forall n \geq 0$ almost surely. We call a distribution π reasonable if under π we a.s. have a reasonable configuration. Since there is no elegant closed form for the necessary and sufficient conditions on the initial fitness distribution to be reasonable, we content ourselves here by giving a couple of sufficient conditions and refer you to Appendix A.1 for a more detailed discussion. π is reasonable if it almost surely sets all fitnesses to 1 apart from the origin, which gets a fitness value less than 1. For finite graphs, we note that uniform product measure is also reasonable. Throughout the rest of this paper, we only consider Bak-Sneppen models with reasonable initial fitness distributions. The term finite Bak-Sneppen model will refer to a Bak-Sneppen model on a finite graph and similarly for infinite Bak-Sneppen model.

2.1.2 A motivating example



The aim of this section is to give some intuition about the model. A heuristic argument is presented to explain how the rules of the model lead to its long term dynamics. In the previous section, you have already seen an important difference between finite and infinite Bak-Sneppen models. This is further elaborated upon in this section. A brief mention is also given to what is meant by

self-organised criticality in the Bak-Sneppen model.

Consider a Bak-Sneppen model on a finite graph with an initial configuration where all fitness values are low. In this case you would expect the update procedure to increase these values, since the updated fitnesses are uniform $(0,1)$ random variables that are on average larger than the original fitnesses. However, if the fitnesses are initially high, you would expect the opposite. Even though the minimal fitness is replaced, the noise effect introduced by replacing the fitnesses of neighbouring vertices would lead to reduced fitnesses on average. It is this sort of dynamic, small values are made larger and large values are made smaller, caused by the combination of updating the smallest value and updating its neighbours that leads to non-trivial limit behaviour of the model. In the case that one selects any vertex rather than the minimal one, uniform product measure is observed as the stationary measure. A similar situation occurs for the Bak-Sneppen model on a complete graph. Φ_n is distributed as uniform product measure and the $\Phi_1, \Phi_2, \ldots$ are independent. The triviality of the Bak-Sneppen model on a complete graph was exploited in [42], see Chapter 3 for more details. If one only updates the minimal vertex (i.e. not its neighbours), then we have a trivial model where the fitnesses converge to 1. This is equivalent to a Bak-Sneppen model on a completely disconnected graph.

Figure 2.1 gives a snap-shot of the Bak-Sneppen model after 300,000 updates. In the initial configuration, all vertices had independent uniform $(0,1)$ fitnesses. We take Λ_{700} as our graph, so the two end points in the diagram are in fact neighbours of each other. Models of this type were those originally considered by Bak and Sneppen. Observe that apart from the vertices near the minimal vertex, most fitnesses appear to be distributed uniformly above a certain value close to $\frac{2}{3}$. In fact, similar pictures are also observed when simulating the Bak-Sneppen model on Λ_{700} when starting from any reasonable initial fitness distribution. Section 2.4 rigorously discusses the limiting behaviour of finite Bak-Sneppen models.

Since the transition rule for the Bak-Sneppen model is local, it is possible to simulate an infinite Bak-Sneppen model up to a given number of time steps. However, there are some crucial differences between finite and infinite Bak-Sneppen models. Note that the independent uniform $(0,1)$ initial fitness distribution, which is the canonical initial fitness distribution for finite Bak-Sneppen models, is not reasonable for an infinite Bak-Sneppen model. It has been common in the literature to consider the limit of Bak-Sneppen models as the graph size tends to infinity rather than looking at Bak-Sneppen models on infinite graphs directly.

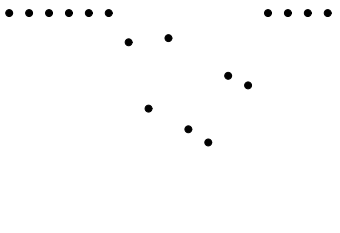
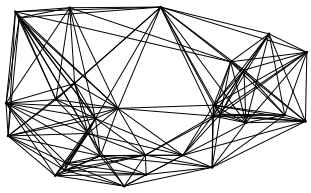


Figure 2.1: A realisation of the Bak-Sneppen model on Λ_{700}.

For example, people have been interested in what happens to the stationary distribution of the Bak-Sneppen model on Λ_N as N tends to infinity. This also invites questions about the differences between looking at a Bak-Sneppen model on an infinite graph and looking at the limit of Bak-Sneppen models on a sequence of finite subgraphs that converge to the infinite graph.

The Bak-Sneppen model is considered self-organised critical because the typical fitness configuration seen in Figure 2.1 is thought to be critical for the model. A critical state is typified by power law distributions. For the Bak-Sneppen model a number of things are proposed to have a power law. I give here just two related examples. Recall from the picture that apart from a small perturbation around the minimal fitness, most fitnesses are situated above some threshold. Let us call this threshold p. It is thought that the number of updates needed so that all values are above this threshold has a power law distribution. Similarly, the number of vertices that need to be updated is thought to have a power law distribution. Both the above claims need some caveats. Of course the number of fitnesses below p initially is important, let us assume that initially there is at most one such vertex. Also finite size effects need to be taken into account in that any power law will only be valid up to a certain value. However, when looking at infinite systems or taking the thermodynamic limit, these finite size cut-offs disappear.



2.1.3 Challenges for the Bak-Sneppen model

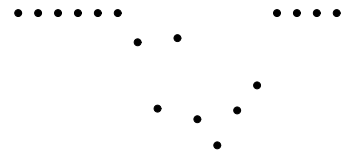
Up to this point, the research agenda has been determined by the interests of the physics commu-

nity. However, since physicists have a different approach and objectives to mathematicians, these are not necessarily close to the interests of mathematicians. One major difference is that physicists tend to assume fundamental properties of the model rather than proving them. Of course these assumptions turn out to be correct in the vast majority of cases, but this sort of approach is not acceptable for a mathematician. One of the goals of this report is to give the model a more rigorous underpinning by proving these basic properties. Another important difference is that many of the objectives of the physicists are beyond the scope of current mathematics. It is natural that by considering heuristic arguments and simulations alongside rigorous methods, many more results can be derived than by purely rigorous mathematics alone. Unfortunately mathematicians have to be a little less ambitious in their targets. For example, much attention has been devoted to the calculation of estimates for various power law exponents of the model. Rigorous calculations of such quantities is typically beyond the reach of current mathematics, even for far simpler models. Even where it is possible, e.g. certain cases of two-dimensional or high dimensional percolation, this requires recent and heavy-weight mathematical methods.

This section contains a list of natural questions that a mathematician might ask about the Bak-Sneppen model. There is a brief motivation of the current state of research which gives rise to some perhaps less natural questions.

1. Stationary Distributions. A natural first step when confronted by a stochastic process is to think about stationary distributions. This is especially relevant for the Bak-Sneppen model since the motivation for looking at the model is the behaviour of the fitnesses in the temporal limit, which can often be described by a stationary distribution. Section 2.4 considers this, proving various results about stationary distributions for finite Bak-Sneppen models. The physics literature has some very precise predictions about the nature of these stationary distributions and proving these predictions is the motivation for the vast majority of mathematical work about the model. Section 2.4 also briefly touches on infinite graphs, where the notion of a stationary distribution is not so relevant.

2. Infinite Volume Limits. Since the Bak-Sneppen model is driven by extremal dynamics (only one vertex can be minimal at any given time), the model is non-Gibbsian. This means that there can be marked differences between infinite Bak-Sneppen models and the thermo-

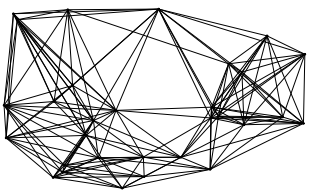


dynamic limit of a sequence of finite Bak-Sneppen models. Only in the special case of $\mathbb{Z}$ and Λ_N has this been properly discussed.

3. Phase Transitions. In common with most self-organised critical models, the Bak-Sneppen model can easily be modified to become a classical model with a tuning parameter [41]. One can then ask all the usual questions related to the behaviour of the model for various values of this tuning parameter. Section 2.3 presents the current theory in this area. The key goal here is to demonstrate the uniqueness of the critical value and also acquire good bounds (or even exact calculations) of the critical value. Questions about the stationary distribution are highly relevant to the stationary distributions of large, but finite, Bak-Sneppen models. There are of course many other more ambitious goals in this area related to investigating the behaviour of the model at criticality. However, such objectives are unlikely to be achieved without further work in the areas listed here.

4. Cluster Generation. When considering the version of the model with a tuning parameter, one can consider the vertices updated by the model as a cluster (in a similar way to the cluster at the origin in a percolation model). The structure of this cluster is unknown for general graphs, although some results have been achieved for the special case of $\mathbb{Z}$ [44]. The structure of these clusters are important for determining the level of similarity between finite Bak-Sneppen models and infinite ones.

5. Random Walk. One can view the location of the minimal fitness as a sort of jump process. It is essentially a simple symmetric random walk with occasional jumps. The behaviour of this random walk is inherently related to the structure of the clusters discussed above. In particular, the recurrence or transience of the random walk is very important to the behaviour of the model in general. Both types of behaviour can occur and depend on the graph [44].



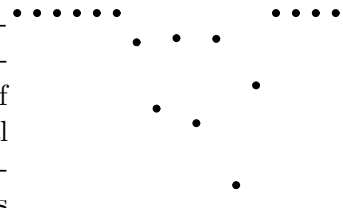
Essentially all the directions for future research mentioned above are variants on the same two themes. Firstly, one wants to gain more insight into the temporal limit behaviour of the model. It is this behaviour that first attracted physicists to study the model. Secondly, it is of great interest to discover how the model behaves on different graphs. This is also closely related to the notion of robustness, namely does the dynamics of the

model change under small modifications? This is again of interest of the wider scientific community, as the plausibility of an SOC description of nature relies on such robustness. There is more on this in the following section.

2.1.4 Variants of the Bak-Sneppen model

It is possible to generalise the Bak-Sneppen model in many ways and people have also considered variants of the model. One of the reasons why people have been so keen to generalise this model comes from its origins as an example of self-organised criticality (SOC). These models were thought to explain the existence of critical systems in nature, by evolving into a critical state without the explicit tuning of a parameter. Think here of percolation, where p has to be tuned to p_c to see critical behaviour. Thus the Bak-Sneppen model was introduced as a robust way of creating critical behaviour and it is natural, therefore, to test this robustness by modifying and generalising the model to see if the new models also display SOC. This paper has already generalised the model by defining the model on graphs other than $\mathbb{Z}$ and Λ_N. Some more dramatic modifications are suggested below. Some of these may be suitable for further mathematical research, but in general are not discussed in this paper. Note that some authors [37, 38] have used the exponential distribution as opposed to the uniform $(0, 1)$ distribution for the marginals of the updated fitnesses. This merely maps the fitnesses of the model from $[0, 1]$ marginals onto $[0, \infty]$ and, therefore, has no real effect on the model.

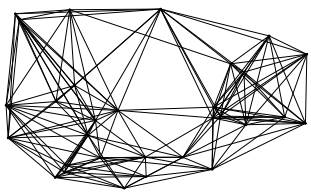
Another natural way to generalise the model, as opposed to changing the graph, is to consider other update rules. One still chooses the minimal (or equivalently maximal) fitness, but the updated fitnesses are no longer independent and uniform $(0, 1)$ random variables. One approach has been to break the equivalence between of the updated fitnesses. For example, one could keep the updated fitnesses of the neighbours random, but update the minimal vertex in such a way that its fitness is always improved by an update. We dub such models modified Bak-Sneppen models in agreement with [25]. Another approach, the deterministic Bak-Sneppen model, is similar except that all updated fitnesses are a deterministic function of their previous values [13, 14]. It appears that if the update function is suitably chaotic, qualitative properties of the model are preserved. Given that the usual treatment of the Bak-Sneppen model in the physical literature has been computer simulation, this



is a natural model to consider. In fact, such a situation can be expressed precisely in the deterministic Bak-Sneppen model setting [27].

It is also possible to more radically alter the Bak-Sneppen model, creating variants that are distinct from the original Bak-Sneppen model as opposed to mere generalisations. Examples of such models are the parallel Bak-Sneppen model [8] and the discrete Bak-Sneppen model [36]. Both these models change the update procedure of the model. In the discrete model, all fitnesses are either 0 or 1 and you choose the minimal site uniformly amongst those with the minimal value. This model is essentially the same as a Bak-Sneppen model where you fix some value p and then pick one vertex with fitness below p to be your 'minimal' vertex. The parallel Bak-Sneppen model is very similar to this, except that all vertices with fitnesses below p are taken to be simultaneously minimal. Such models are closely related to discrete contact processes [26]. Another generalisation of the model is to replace the single fitness of each vertex by a collection of values called *traits* [7]. The overall fitness of each vertex can then be thought of as a function of these traits with updates being driven by the location of the minimal trait. In [16], extinction and branching is added to the model.

One important variant of the Bak-Sneppen model is its mean field approximation [6]. It is common in statistical physics to analyse a model by considering a simplification of the model obtained by ignoring correlations. In the Bak-Sneppen model, correlations between fitness values stem from the fact that one updates the minimal vertex and its neighbours. Therefore, the mean field approximation for this model takes the form of ignoring the graph structure. For a finite regular graph with common vertex degree Δ, this is done by updating, along with the minimal vertex, Δ other vertices chosen uniformly at random from the other vertices of the graph. Another way of thinking about this is replacing our graph with the complete graph of the same order and then taking the update rule that you update not all the neighbours of the minimal fitness, but Δ neighbours chosen at random. For an infinite regular graph with common vertex degree Δ and an initial configuration with only finitely many non-trivial fitnesses, this takes the form of updating the minimal fitness along with Δ previously unconsidered vertices with initial fitness 1. There are very strong similarities between this and a Galton-Watson branching process. Typically correlations are small when the dimension of the model is high. In particular, it is typically the case that certain quantities of interest about the original model in high dimensions are exactly those of the mean field approximation.



Therefore it is natural to consider the mean field approximation of the Bak-Sneppen model in order to gain insight about the Bak-Sneppen model in high dimensions. For the Bak-Sneppen model on $\mathbb{Z}^d$ there has some debate about the dimension at which the model assumes its mean field values [15, 17]. See Appendix A.2 for a brief mathematical treatment of the mean-field model.

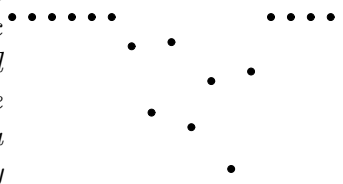
2.2 The Avalanche Decomposition

The main technique for analysing the BS model has been to decompose the model into a series of *avalanches*. An avalanche from a threshold p, referred to as a p-avalanche, is said to occur between times n and $n+m$ if at time n all the fitnesses are equal to or greater than p with at most one vertex where equality holds, and time $n+m$ is the first time after n at which all fitnesses are larger than p. The vertex with the minimal fitness at time n is called the *origin* of the avalanche. Note that by this definition a p-avalanche occurs in the time interval $[n, n+1]$ if at times n and $n+1$ there are no fitnesses less than p.

One can view the Bak-Sneppen model as a sequence of avalanches. At the end of each avalanche, a new avalanche is started with its origin being the minimal vertex. Since the definition states that all fitnesses must be equal to or greater than the threshold, one can choose the threshold to be any value equal to or smaller than the minimal fitness. A common approach is to always choose the maximum possible threshold. Viewing the Bak-Sneppen model in such a way is called the *maximal avalanche decomposition.*

A p-avalanche can be considered as a stochastic process in its own right. The key feature of the origin is that it has the minimal fitness (as it will be updated immediately). Hence, we can consider its fitness to be any value, as long as this value is minimal. Vertices with fitness below the threshold are called *active*, others are called *inactive*. Note that the exact fitness value of an inactive vertex is irrelevant for the avalanche, since this value can never be minimal during the avalanche. This motivates the following formal definition of an avalanche.

Definition 2.3. *A p-avalanche with origin v on a graph G (with vertex set $V(G)$) is a stochastic process with state space $\{[0,p]^A, A \subset V(G)\}$ and initial state $p^{\{v\}}$. The process follows the update rules of the Bak-Sneppen model. Any vertex with a fitness smaller than or equal to p is included. Any vertex with a fitness larger than p is not included.*



The process terminates when it is the empty set.

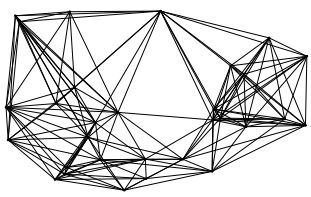
It is useful to define a number of quantities associated with an avalanche. The range set of an avalanche is the collection of vertices updated during the avalanche and the range is the size of this set. The duration is naturally the number of updating steps. An avalanche is called infinite if it has infinite duration. Since the Bak-Sneppen model can be considered as a series of avalanches, it is natural to study the behaviour of individual avalanches. Note that by considering avalanches one has recovered a tuning parameter for the Bak-Sneppen model. This simplifies things somewhat, whilst still having relevance for the original model, since the Bak-Sneppen model is merely a sequence of avalanches.

To introduce the reader to the typical way of generating an avalanche, we prove some trivial results about avalanches. We begin by showing that p-avalanches on finite graphs are finite.

Proposition 2.4. *On a finite graph, p-avalanches are finite almost surely for all $p < 1$. Furthermore, the expected duration of such avalanches is also finite.*

Proof: Fix the graph G and threshold p. Let Δ denote the maximal degree of the graph and N be the number of vertices of the graph. Consider a sequence of independent uniform $(0, 1)$ random variables, $\{U_i\}_{i\geq 1}$. This sequence is used to realise the p-avalanche on G. To generate the new fitnesses of updated vertices, we draw the next value from the sequence $\{U_i\}_{i\geq 1}$. At each time step, a number of vertices are simultaneously updated. We arbitrarily designate an order to these vertices when drawing the new fitnesses from our sequence.

We call a block $U_k, U_{k+1}, \ldots, U_{k+\Delta(n+1)}$ good if $U_{k+j} > p$ for all $0 \leq j \leq \Delta(n+1)$. Note that there almost surely exists a good block for some k. There are now two cases. If every vertex is updated to a random variable from this block, then all fitnesses are greater than p and the avalanche stops. If there is some vertex that does not receive a fitness value from one of the random variables in the block, then at some point one of the values from this block becomes minimal. However, this means that the minimal fitness is greater than p and the avalanche stops. Therefore a good block is a sufficient condition for the p-avalanche to end. Since a good block occurs almost surely, the avalanche is finite almost surely.



The proof of the finite expected duration is similar. Now partition the sequence into non-overlapping blocks of length $\Delta(n+1)$. Taking the same notion of a good block from above, we have that

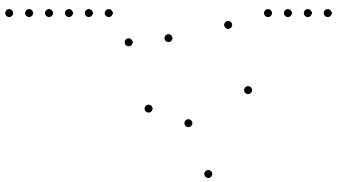
the probability of one of our blocks being good, p_G is positive. Since the blocks are non-overlapping, the events of them being good are independent. Thus we can bound from above the expected duration of the avalanche by the expected value of a geometric distribution with parameter p_G. This is merely p_G^{-1} and therefore finite. □

The following domination result follows from using the avalanche construction above to simultaneously generate two avalanches of differing thresholds.

Proposition 2.5. *If $p > p'$, then there exits a coupling of a p-avalanche and a p'-avalanche on the same graph with the same origin such that the two avalanches are identical up to the termination of the p'-avalanche.*

Proof: Recall from the proof of Proposition 2.4 the method of generating an avalanche from a sequence independent uniform $(0, 1)$ random variables, $\{U_i\}_{i\geq 1}$. Take two identical copies of our graph G and generate two avalanches, one with threshold p, the other with threshold p' using the same sequence of sequence of uniform random variables and the same arbitrary ordering of the vertices when drawing the new fitnesses from the sequence. This coupling has the property that the fitnesses in both avalanches are identical. Therefore, the p-avalanche stopping implies that the p'-avalanche must also stop (if it has not finished already). Furthermore, the behaviour of the avalanches are identical up to the point that the p'-avalanche stops. This yields the desired results. □

Proposition 2.4 concerns itself with avalanches with a fixed threshold on finite graphs. However in some circumstances, it is more appropriate to consider avalanches on finite graphs with a random threshold. This situation is dealt with in Chapter 3. Consider for example the first avalanche in the maximal decomposition of a Bak-Sneppen model, where the threshold is some random variable with law determined by the initial fitness distribution. 1-avalanches are by definition infinite and so it comes as no surprise to find that it is the tail of the threshold distribution as the threshold tends to 1 that determines whether the expected duration of the avalanche is infinite. It is perhaps more surprising that in the most natural settings, the expected avalanche durations are infinite. Examples of cases where you have infinite expected duration on the circular graph include the first maximal avalanche when the initial condition of all fitnesses is uniform $(0, 1)$ product measure and the maximal avalanche decomposi-



tion starting from any initial fitness configuration [42].

2.2.1 Avalanche Critical Values

Recall from Definition 2.3 that the behaviour of an avalanche doesn't depend on the exact fitness values. Given a graph, G, it is only required to know the threshold and origin of the avalanche. We define the following quantities. For a p-avalanche on a graph G with origin v, we denote its range set by $\xi_{G,v}(p)$ and duration by $\eta_{G,v}(p)$. We also have the range, $r_{G,v}(p) = |\xi_{G,v}(p)|$. An important special case is when the range of the avalanche is $|G|$, i.e. for finite graphs this means that the range set is the entire vertex set of the graph.

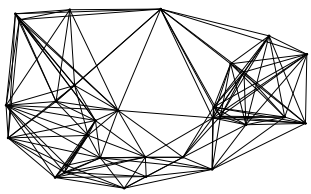
Definition 2.6. *A spanning avalanche on a graph, G, is an avalanche that updates every vertex of G.*

To help analyse avalanches, we introduce the following mean values.

$$\begin{aligned} R_{G,v}(p) &= E(r_{G,v}(p)), \\ D_{G,v}(p) &= E(\eta_{G,v}(p)), \\ P_{G,v}(p) &= P(r_{G,v}(p) = |G|). \end{aligned}$$

Thus $R_{G,v}(p)$ is the expected range and $D_{G,v}(p)$ is the expected duration of the avalanche. $P_{G,v}(p)$ is the probability of a spanning avalanche if G is finite and the probability of an infinite avalanche otherwise. It is an immediate consequence of Proposition 2.5 that $R_{G,v}(p)$, $D_{G,v}(p)$ and $P_{G,v}(p)$ are non-decreasing in p. Also note that when G is transitive it is not necessary to specify the origin, as this does not affect the mean values.

Meester and Znamenski [38] considered the special case when G is the circular graph. A number of their results follow from similar arguments to those above, using what can be considered as the *natural coupling*. However, in some cases this technique was not sufficient and something called the *self-similar graphical representation* was necessary. Their results are stated here without proof.



Proposition 2.7. *$R_{\Lambda_N}(p)$ is non-decreasing in N, $P_{\Lambda_N}(p)$ is non-increasing in N and $D_{\Lambda_N}(p)$ is non-decreasing in N.*

We now turn our attention to infinite graphs. Recall that an infinite avalanche was an avalanche with infinite duration. However, it is an almost immediate consequence of Proposition 2.4, that an infinite avalanche (on an infinite graph) has infinite

duration almost surely. Exploiting monotonicity, Proposition 2.5, we define the following critical values.

$$\begin{aligned} p_c^p(G) &= \inf\{p : P_G(p) > 0\} \\ p_c^r(G) &= \inf\{p : R_G(p) = \infty\} \\ p_c^d(G) &= \inf\{p : D_G(p) = \infty\} \end{aligned}$$

Again the origin of the avalanche is suppressed, because a p-avalanche with origin v ends up in a state identical to the beginning of a p-avalanche with origin v' with positive probability. Therefore the critical values are independent of the choice of origin.

The three critical thresholds have a natural order. The possibility of an infinite avalanche implies an infinite expected range. Restricting ourselves further to graphs of bounded degree, we have that $D_G(p) \geq (\Delta(G)-1)R_G(p) - \Delta(G)$, where $\Delta(G)$ is the maximal degree of the graph. Thus,

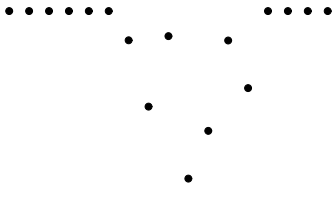
$$p_c^d(G) \leq p_c^r(G) \leq p_c^p(G), \text{ when } \Delta(G) < \infty.$$

Section 2.3 summarises results known about these critical values and Section 2.4 explains the relationship between these critical values and the behaviour of finite Bak-Sneppen models. When confronted by a number of different definitions for critical values it is natural to ask if these definitions are equivalent. Typically the equivalence or otherwise of critical values gives significant insight into the behaviour of the model itself. The importance of this can be seen in Section 2.4. So far the only result in this direction is that $p_c^d(\mathbb{Z}) = p_c^r(\mathbb{Z})$ [38]. This is a consequence of the following differential equation.

Proposition 2.8. *If $R_{\mathbb{Z}}(p) < \infty$, then $D_{\mathbb{Z}}(p)$ is differentiable and*

$$\frac{dD_{\mathbb{Z}}(p)}{dp} = \frac{D_{\mathbb{Z}}(p)R_{\mathbb{Z}}(p)}{1-p} \text{ if } R_{\mathbb{Z}}(p) < \infty.$$

At this point, it is worth making a short note about the behaviour of the limits of the avalanche mean values as the graph size goes to infinity. It has been shown [38] that for any $p > 0$ and any $N \geq 3$,



$$\begin{aligned} R_{\Lambda_N}(p) \leq R_{\mathbb{Z}}(p), &\quad \lim_{N\to\infty} R_{\Lambda_N}(p) = R_{\mathbb{Z}}(p), \\ D_{\Lambda_N}(p) \leq D_{\mathbb{Z}}(p), &\quad \lim_{N\to\infty} D_{\Lambda_N}(p) = D_{\mathbb{Z}}(p), \\ P_{\Lambda_N}(p) \geq P_{\mathbb{Z}}(p), &\quad \lim_{N\to\infty} P_{\Lambda_N}(p) = P_{\mathbb{Z}}(p). \end{aligned}$$

However, these results have only been shown to apply for $\mathbb{Z}$, which is a special case. It has been shown [44] that an infinite avalanche on $\mathbb{Z}$ updates every vertex. However, on other graphs it is possible to be infinite without updating all vertices. This has been demonstrated for a rapidly branching tree, but is also thought to be possible for more standard graphs. Thus, it is unreasonable to expect the limiting results given above to hold on such graphs.

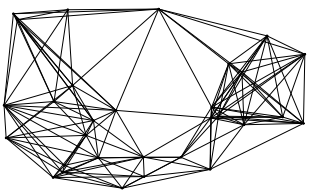
Since the primary focus of physicists has been to look at the behaviour of large but finite systems, the above approach also makes sense for graphs other than $\mathbb{Z}$. However, the results above have only been shown to apply to BS models on $\mathbb{Z}$. An alternative approach where the limits and domination hold by definition is to take finite boxes of an infinite graph and consider the limits of the range and duration with a cutoff when the edge of the box is reached. Similarly one could consider the probability of reaching the edge of the box in place of the probability of a spanning avalanche. Since infinite avalanches on some graphs do not update all vertices, it seems likely that the probability of a spanning avalanche on a large box is not going to be closely related to the critical value of the infinite graph.

2.2.2 The Locking Threshold Representation

An alternative way of viewing the Bak-Sneppen model called the *locking threshold representation* was introduced in [38]. This was originally described for the Bak-Sneppen model with exponentially distributed fitnesses. We translate this here into the terminology of the usual setting with uniform marginals.

Consider a Bak-Sneppen model on a finite graph G where all the initial fitnesses are independent uniform $(0,1)$ random variables. Let N be the number of vertices of G, $v_1, \ldots, v_N$ denote the vertices of G and denote the fitnesses of these vertices at time n by $X_1(n), X_2(n), \ldots, X_N(n)$. For brevity we shall denote the collection of fitnesses by $\mathcal{X}(n)$. Let $F_N(n,.)$ denote the joint distribution function of the fitnesses at time n. Hence,

$$F_N(n, x_1, \ldots, x_N) = \mathbb{P}(X_1(n) < x_1, \ldots, X_N(n) < x_N).$$



We now define the locking thresholds $\mathcal{Y}(n) = (Y_1(n), \ldots, Y_N(n))$ and show that given the locking thresholds, the fitnesses $\mathcal{X}(n)$ are independent and uniformly distributed above their locking thresholds. For example, $X_i(n)$ is uniform $(Y_i(n), 1)$.

Let $Y_i(0) = 0$ for $1 \leq i \leq N$. Now suppose that at some time n, we have a collection of fitnesses

$\mathcal{X}(n)$ and locking thresholds $\mathcal{Y}_i(n)$. Let $v_{I(n)}$ be the vertex with minimal fitness at time n with value $X_{I(n)}(n)$. Then,

$$Y_i(n+1) = \begin{cases} \max\left(Y_i(n), X_{I(n)}(n)\right), & \text{for } v_i \in G \backslash \Gamma^\star(v_{I(n)}), \\ 0, & \text{for } v_i \in \Gamma^\star(v_{I(n)}). \end{cases} \tag{2.1}$$

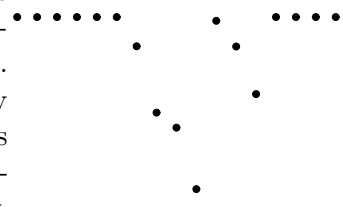
Proposition 2.9. *Given their locking thresholds, the fitnesses of a finite Bak-Sneppen model (with i.i.d. uniform* $(0,1)$ *initial distribution) are independent and uniformly distributed between their locking thresholds and* 1*.*

Proof: We proceed by induction. Initially all fitnesses are independent and uniform $(0,1)$, agreeing with our $Y_i(0) = 0$ for $1 \leq i \leq N$. Now consider an arbitrary time N where the fitnesses are uniformly distributed between their locking thresholds and 1. We now have two cases. If a vertex is minimal or a neighbour of the minimal vertex, then it will be updated receiving a new independent uniform $(0,1)$ distributed fitness agreeing with (2.1). Otherwise recall by assumption that the vertex is distributed uniformly between its locking threshold and 1. So consider a random variable U, which is distributed uniformly on $(u,1)$. Let $\hat{U}$ denote U conditioned on $U > m$, where m denotes the value of the minimal fitness. Then by the properties of the uniform distribution, $\hat{U}$ is distributed uniformly on $(\max(u,m),1)$. □

Figure 2.2 gives a picture of what the locking threshold representation of the Bak-Sneppen model typically looks like. We have used the same conditions as Figure 2.1, but Figure 2.2 shows a different realisation of the model. Note that the majority of the locking thresholds take the same value, with an area of lower thresholds around the location of the minimum. This is equivalent to the observation that the minimal fitnesses in Figure 2.1 seem to be uniformly distributed above some threshold.

2.2.3 The Forgetful Avalanche Construction

The *forgetful avalanche construction* was first used in [42], see Chapter 3 and first mentioned as such in [43], see Chapter 4. Its primary strength is that it permits a number of couplings between the Bak-Sneppen model and other more basic models. It uses the locking thresholds representation to simultaneously generate a collection of realisations of the Bak-Sneppen model. Typically it is easier to pair off in a meaningful way these collections with realisations of other models (e.g. site percolation) than by considering individual realisations of the Bak-Sneppen model directly.



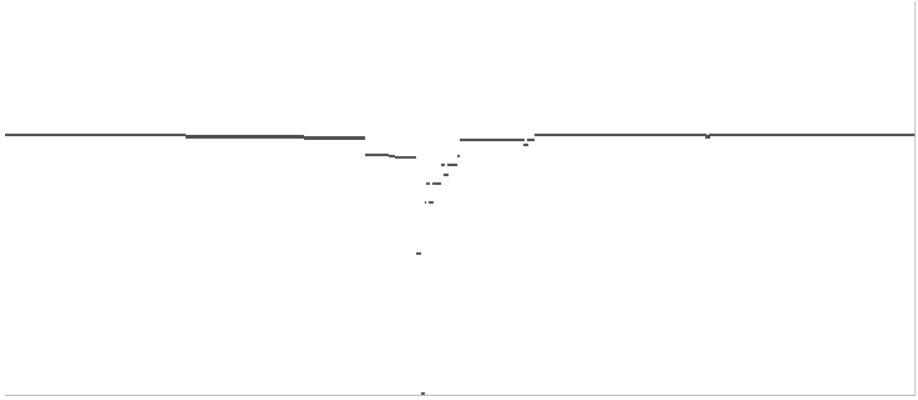


Figure 2.2: The locking thresholds representation of the Bak-Sneppen model on Λ_{700}.

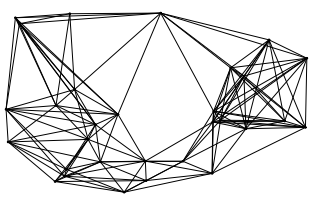
Readers are directed to Chapter 4 for a description of this approach in terms of the original model.

Before defining forgetful avalanches in their own right, we need the following notation. For $x \in (0,1)$, we denote the uniform $(x,1)$ distribution by f_x. The point mass at 1 is denoted by f_1.

Definition 2.10. *A forgetful p-avalanche, η_n on a connected locally finite graph, G, with origin o is a discrete time stochastic process with state space $[0,1]^{V(G)}$. Initially, $\eta_0(v) = 0$ if $v = o$ and $\eta_0(v) = 1$ otherwise. η_{n+1} is generated from η_n by the following procedure. For each $v \in v(G)$ such that $\eta_n(v) < 1$, an independent $f_{\eta_n(v)}$ random variable is realised. Let M denote the minimum of these values and v_M its associated vertex. Then,*

$$\eta_{n+1} = \begin{cases} 0 & \text{for } v \in \Gamma^\star(v_M) \\ M \wedge \eta_n(v) & \text{otherwise.} \end{cases}$$

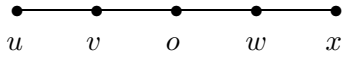
The connection between this and the original Bak-Sneppen model should be clear from the locking threshold representation. A forgetful avalanche, so called because the exact fitness values are forgotten, is merely the Bak-Sneppen model viewed from the perspective of its locking thresholds. For example, replacing the $\eta_n(v)$ with independent $f_{\eta_n(v)}$ random variables recovers the original model. However, the forgetful avalanche construction represents all possible such realisations simultaneously and lets all avalanches with the same sequence of



minimal vertices and fitnesses be considered together.

Vertices	u	v	o	w	x
η_0	1	1	0	1	1
RVs			0.5		
η_1	1	0	0	0	1
RVs		0.2	0.3	0.7	
η_2	0	0	0	0.2	1
RVs	0.9	0.5	0.8	0.3	
η_3	0.3	0.3	0	0	0
RVs	0.8	0.7	0.1	0.4	0.3
η_4	0.3	0	0	0	0.1

The thresholds and fitnesses

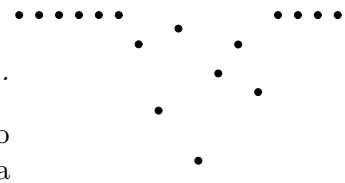


The graph structure

Figure 2.3: An example of a forgetful avalanche

2.3 Bounds for avalanche critical values

In Section 2.2, three avalanche critical values were defined. In this section, we aim to show that these values are non-trivial, i.e. that avalanches exhibit a phase transition, on a wide range of graphs. Although it is in general not possible to calculate the exact values of these critical values, some bounds can be calculated. We begin first with a general approach for showing that $p_c^p(G) < 1$. This first appeared in [37] and is perhaps more clearly presented in [39]. It should be noted that these results were originally phrased in terms of bounds for the limiting stationary distribution of increasing sequences of graphs with exponential fitnesses. This connection will become clear in Section 2.4. The conditions for their original theorem are rather technical and are to do with the regularity and growth properties of the graph. To avoid such technicalities, the theorem is presented here is a weaker form.



Theorem 2.11. *If G is transitive then $p_c^p(G) < 1$.*

It is also possible from the method of proof to calculate an explicit upper bound. However, in a similar way to the bounds obtained by a Peierls

type argument, this bound increases as you increase the dimension, whilst the actual critical values are decreasing. However, in the special case of the line this bound is the best available and worth stating.

Corollary 2.12. $p_c^p(\mathbb{Z}) \leq 1 - e^{-68}$.

In decimal notation, this bound is roughly a decimal point followed by thirty nines. The same authors also calculated a bound for $p_c^r(\mathbb{Z})$.

Proposition 2.13. $p_c^r(\mathbb{Z}) \leq 1 - e^{-2\ln 2} = \frac{3}{4}$

An alternative approach to bounding the avalanche critical values was used in [43]. The lower bound is a standard and well-known result. We let $p_c^{BP}(\Delta + 1)$ denote the critical value of a branching process with binomial $(\Delta + 1, p)$ offspring distribution.

Theorem 2.14. *Let G be a regular graph with common vertex degree Δ, then*

$$p_c^{BP}(\Delta + 1) \leq p_c^b(G),$$

for $b \in \{d, r, p\}$.

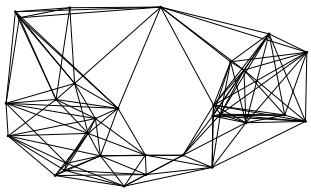
The proof of Theorem 2.14 can be found in Chapter 4. This result can also be extended to other graphs (such as spherically symmetric trees) by considering multi-type branching processes. In the case of a spherically symmetric tree, the different particle types would correspond to the different generations of the tree. The lower bound in Theorem 2.14 is easy to calculate, but it is also possible from the method of proof to consider any graph of bounded degree.

Corollary 2.15. *If G is a graph with maximal degree Δ, then*

$$\frac{1}{\Delta + 1} \leq p_c^b(G),$$

for $b \in \{d, r, p\}$.

The upper bound is more subtle than the lower bound and requires the use of the forgetful avalanche construction.



Theorem 2.16. *For any G we have*

$$p_c^b(G) \leq p_c^{site}(G),$$

where $p_c^{site}(G)$ denotes the critical value of independent site percolation on G and $b \in \{d, r, p\}$.

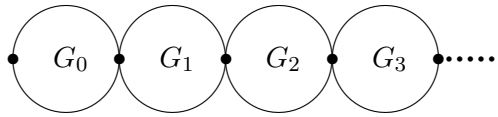


Figure 2.4: The $\{G_i\}_{i\geq 1}$ are graphs such that $\Delta(G_i) \to \infty$ but $|G_i| < \infty \ \forall i$.

Since independent site percolation is a simpler model than the Bak-Sneppen model, this inequality yields useful bounds in a number of cases.

Corollary 2.17. *Let T_Δ denote a regular tree with common degree Δ, then*

$$p_c^p(T_\Delta) \leq \frac{1}{\Delta - 1}.$$

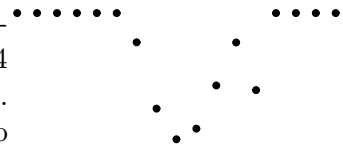
The following upper bound comes from Hara and Slade [35].

Corollary 2.18. *As $d \to \infty$,*

$$p_c^p(\mathbb{Z}^d) \leq \frac{1}{2d} + \frac{1}{(2d)^2} + O(d^{-3}).$$

Note that Theorem 2.11 can be extended to apply to the 'smallest' infinite graph, $\mathbb{Z}^+$. Morally, one would expect the critical value of any other infinite graph to be smaller than this. However, this monotonicity is unproven and it is possible to find graphs which satisfy neither the conditions of Theorem 2.11 nor gain a non-trivial bound from Theorem 2.16. Consider for example a graph of unbounded degree shown in Figure 2.3. Clearly site percolation on such a graph only occurs when $p = 1$, for the same reasons as for $\mathbb{Z}$. However the graph has unbounded degree, so Theorem 2.11 is also of no use.

Conjecture 2.19. *For any infinite graph G, $p_c^p(G) < 1$.*



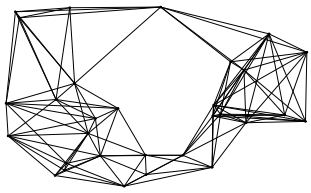
Note that the opposite claim concerning non-triviality of $p_c^p(G)$ around 0 is false. Theorem 2.14 shows that $p_c^p(G) > 0$ for a wide range of graphs. However, graphs with explosive vertex degree do not have a degree distribution and are not included by the conditions of the theorem and $p_c^p(G) = 0$ for

some G. For example, in [44] a rooted tree is considered. It is shown that a 1-avalanche on his graph with origin at the root always jumps forward to the next generation with positive probability. In other words, the minimal vertex at time n is a member of generation n. Fix $\epsilon > 0$ and claim that $p_c(T) > \epsilon$. Now decompose our 1-avalanche into a sequence of ϵ-avalanches which are finite almost surely by assumption. With probability 1 there are infinitely many times when all fitnesses are greater than ϵ (each time an ϵ-avalanche finishes). Thus, this still must happen infinitely often when conditioning on the 1-avalanche having its minimal vertex at time n in generation n for all n. However, the event that the parent of the minimal vertex (i.e. the neighbour of the minimal vertex in the previous generation) never receives a fitness less than ϵ has probability 0. Therefore at some point a parent will have an updated fitness less than ϵ. This vertex is never updated again by the assumption that the minimal fitness always moves to the next generation. However, this contradicts the fact that all fitnesses are greater than ϵ infinitely often.

2.4 Stationary Distributions

For a finite Bak-Sneppen model it is natural to think about the existence of stationary distributions and their nature. Computer simulations strongly suggest that the long term behaviour of finite Bak-Sneppen models does not depend on the initial fitnesses. In this section we prove that for any finite graph, the Bak-Sneppen model converges weakly to a unique stationary distribution. For the special cases of Λ_N, this has long been assumed in the literature. We present here a coupling proof of the convergence to a unique stationary distribution. It should be noted that this result can also be derived via the standard tools for Markov chains on uncountable state spaces laid out, for example, in [32]. Interestingly this abstract Markov approach can be realised using the same constructions presented below.

Having established the uniqueness of the stationary distribution, some properties of this distribution are derived. Of particular interest is the asymptotic behaviour of the stationary distribution as the size of the graph goes to infinity. This will be in stark contrast to the end of this section, where the limiting behaviour of the Bak-Sneppen model on infinite graphs is discussed.



2.4.1 Uniqueness by coupling

In this section we prove the following theorem, the proof of which is concluded by the proof of Propo-

sition 2.23.

Theorem 2.20. *The Bak-Sneppen model on a finite graph G has a unique stationary distribution, π_G. Furthermore, $\Phi_n \xrightarrow{\mathrm{d}} \pi_G$ from all reasonable Φ_0.*

We begin by showing existence of stationary distributions for finite Bak-Sneppen models.

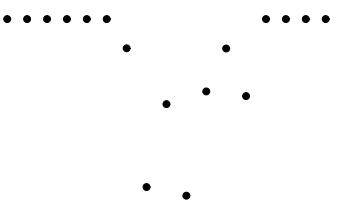
Lemma 2.21. *Any finite Bak-Sneppen model has at least one stationary distribution.*

Proof: Recall that the Bak-Sneppen model is only defined on $\Omega_G \subset [0,1]^{V(G)}$. In the finite case, Ω_G is merely $[0,1]^{V(G)}$ with all configurations where there is a tie for the minimal fitness removed. Observe that although the space $[0,1]^{V(G)}$ is compact, Ω_G is not. There is a classical result from ergodic theory that states that a stochastic process on a compact state space has at least one stationary distribution. An elegant presentation of the proof of this result can be found in [31], where the result is traced back to a paper from 1937 [34]. In order to apply this result to the Bak-Sneppen model, it is necessary to first extend the model from Ω_G to $[0,1]^{V(G)}$. By extend, I mean creating a model which has the same law as the Bak-Sneppen model on Ω_G, but that is also well-defined on $[0,1]^{V(G)}\backslash\Omega_G$. There is considerable freedom in how to do this, but for our purposes it is important to extend the model in such a way that we can't get trapped on $[0,1]^{V(G)}\backslash\Omega_G$. We consider the vertex set $V(G)$ to be ordered and in the case of a tie for the minimal fitness, the 'minimal' vertex is the vertex with the lowest index amongst those of minimal fitness.

It is clear that the extended model has a stationary distribution, since its state space is compact. We proceed by showing that any stationary distribution for the extended model is also a stationary distribution for the BS model. Since the two models are identical on Ω_G it is sufficient to show that $\pi_G([0,1]^{V(G)}\backslash\Omega_G) = 0$, where π_G is an arbitrary stationary distribution of the extended model. Let

$$A_i = \{\omega \in [0,1]^{V(G)} : \#\{j : \omega_j = \omega_k \text{ for some } k \neq j\} = i\}.$$

Thus the set A_i contains all configurations with exactly i tied fitnesses, so $A_1 = \emptyset$. Clearly $[0,1]^{V(G)}\backslash\Omega_G \subset \cup_{i>1} A_i$. Let $B_i = A_i \cap [0,1]^{V(G)}\backslash\Omega_G$, so $\cup B_i = [0,1]^{V(G)}\backslash\Omega_G$. We shall demonstrate that $\pi_G(B_i) = 0$ for any $i > 0$. The update rule of the extended model creates almost surely no new tied values and $\forall k > 1$ maps each configuration in B_k

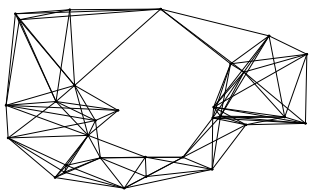


to some B_i where $i < k$. Note that $B_n = \emptyset$ for $n > |G|$. Thus by the properties of a stationary distribution, $\pi_G(B_{|G|}) = 0$. Applying this inductively, we see that $\pi_G(B_i) = 0$ for any $i > 0$. Therefore any stationary distribution, π_G, on the extended model satisfies $\pi_G([0,1]^{V(G)} \backslash \Omega_G) = 0$ and is also a stationary distribution for the Bak-Sneppen model. □

Having used an extension of the Bak-Sneppen model to prove the existence of a stationary distribution, we return to considering the original model. To prove uniqueness and convergence we use a coupling approach. We take two copies of our graph G, G_1 and G_2 , and run Bak-Sneppen models, Φ^1 and Φ^2 on both copies. $\Phi^1_0 \sim F_1$ and $\Phi^2_0 \sim F_2$. We couple the two models in such a way that there a.s. exists a random time $T < \infty$ such that $\Phi^1_n = \Phi^2_n$ for all $n \geq T$. If F_1 is a stationary distribution then $\Phi_n \sim F \ \forall n$, therefore $\Phi^2_n \to F_1$. Hence F_1 is the unique stationary distribution and the model converges weakly to this distribution. It now suffices to construct such a coupling.

Before introducing this coupling, we need to define a special method for realising a Bak-Sneppen model on a given finite graph. Let $|G| = N$ and Δ denote the maximal degree of G. We denote the vertices of G by $v_1, v_2, \ldots, v_N$ and consider this as an ordering of the vertices. Let $\{U_i\}_{i \geq 1}$ be the usual i.i.d. sequence of $U(0,1)$ random variables. Consider a cycle $K = (x_1, x_2, \ldots, x_M)$ such that $x_i \in \{v_1, v_2, \ldots, v_N\}$, $x_i \sim x_{i+1} \ \forall i < M$, $x_1 \sim x_M$ and for each $v_j \ \exists i$ such that $x_i = v_j$. In words, K is a cycle that visits every vertex of G at least once. Since G is connected, such a cycle always exists. Consider a minimal such cycle.

The cycle K can be used together with the sequence $\{U_i\}_{i \geq 1}$ to update the model. Denote the minimal fitness at time i by I_n and consider the minimal vertex as a point in the cycle, $I_n = x_i$ for some i. The fitnesses at time $n+1$ are drawn from the set $\{U_{(\Delta+1)n+1}, U_{(\Delta+1)n+2}, \ldots, U_{(\Delta+1)(n+1)}\}$. The order of allocating these values to the vertices of $\Gamma^\star(x_i)$ is determined by K and the original ordering of the vertices. The new fitness of x_{i+1}, $\Phi_{n+1}(x_{i+1}) = U_{(\Delta+1)n+1}$ with the rest of the fitnesses being allocated in increasing index order. Note that if degree of I_n is less than Δ, then only the first few random variables are used and the remaining ones are discarded. I_{n+1} is found by moving along the cycle from x_i until the minimal vertex is encountered.



We need the following preliminary lemma.

Lemma 2.22. *For any $0 < p < 1$ and any finite graph G, there exists a $q > 0$ such that $P_{G,v}(p) \geq q$ $\forall v \in V(G)$.*

Proof: Recall the method of generating the model according to a cycle given above. We call the block $\{U_i\}_{1\leq i\leq m(\Delta+1)}$ good if

$$U_i = \begin{cases} < p & \text{for } i = 1 \mod \Delta + 1 \\ > p & \text{otherwise.} \end{cases}$$

If we generate the avalanche from a good block then the location of the minimal fitness walks around the entire vertex set and always has value less than p. The probability of a good sequence, q, is clearly independent of the origin of the avalanche and is positive forall $0 < p < 1$. Therefore $P_{G,v}(p) \geq q \ \forall v \in V(G)$. □

The coupling approach is to allow Φ^1 and Φ^2 to evolve independently until the minimal vertices are at the same location in both graphs and all other vertices have fitnesses greater than p. We then couple the two models by updating them using the same cycle and sequence of uniform random variables. Then by Lemma 2.22, there is a uniformly bounded probability that the subsequent avalanche is spanning (in both graphs). In this case the fitnesses of both graphs are identical and by continuing to use the same cycle and random variables, they remain so forever. The proof of Theorem 2.20 is concluded by the following proposition.

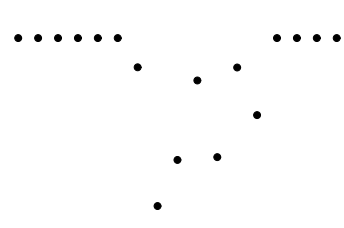
Proposition 2.23. *Let Φ^1 and Φ^2 be two independent Bak-Sneppen models with arbitrary initial conditions. Fix $0 < p < 1$. Let $\tau_1, \tau_2, \ldots$ be a sequence of stopping times such that*

$$\tau_{i+1} = \inf_{n>\tau_i} \{n : I_n^1 = I_n^2 \text{ and } \Phi_n^b(v_i) > p \ \forall b \in \{1,2\} \text{ and } 1 \leq i \leq N\}$$

where $\tau_0 = 0$. Then $\{\tau_i\}_{i\geq 0}$ is an infinite sequence and $\mathbb{E}(\tau_i - \tau_{i-1}) < \infty \ \forall i \geq 1$.

Proof: Let the independent sequences $\{U_i\}_{i\geq 1}$ and $\{V_i\}_{i\geq 1}$ define Φ^1 and Φ^2 respectively with the same cycle K for both graphs. By the independence of $\{U_i\}_{i\geq 1}$ and $\{V_i\}_{i\geq 1}$, Φ^1 and Φ^2 are independent. We divide both sequences into blocks of length $(\Delta + 1)(N + M)$. We call the block $\{U_i\}_{1\leq i\leq(\Delta+1)(N+M)}$ good if

$$U_i \begin{cases} > p & \text{if } i \leq (\Delta+1)N \\ < p & \text{if } (\Delta+1)N < i < (\Delta+1)(N+M-1) \\ & \quad \text{and } i = 1 \mod \Delta + 1 \\ > p & \text{if } (\Delta+1)N < i \leq (\Delta+1)(N+M) \\ & \quad \text{and } i \neq 1 \mod \Delta + 1 \\ > p & \text{if } i = (\Delta+1)(N+M-1)+1. \end{cases}$$



The effect of a good block is to leave all fitnesses of the model independent and uniformly distributed on $[p, 1]$. Note that this construction negates the usual aging effect, i.e. that older fitnesses are stochastically larger. The first $(\Delta + 1)N$ random variables ensure that all fitnesses are greater than p. Therefore, the rest of the good block walks the minimal fitness along K updating all vertices. At the end of a good block, the probability that any vertex is minimal is $1/N$.

Assume that the i-th blocks of both $\{U_i\}_{i\geq 1}$ and $\{V_i\}_{i\geq 1}$ are good. Then the minimal vertices both have fitnesses greater than p and are in the same location in both graphs with probability $1/N$. Using these blocks we can then construct a subsequence of $\{\tau_i\}_{i\geq 1}$ with the desired properties. It is clear that the probability of a good block is positive and the probability that a pair of blocks is good is just the square of this, by independence. Therefore by Borel-Cantelli, we recover an infinite sequence of pairs of good blocks that leave the minimal fitness in the same place on both graphs with all fitnesses greater than p. Furthermore, $\mathbb{E}(\tau_1) < \infty$ since it is the mean of a geometric distribution. Since this holds for arbitrary initial conditions, it is immediate that $\mathbb{E}(\tau_i - \tau_{i-1}) < \infty \ \forall i \geq 1$ also. □

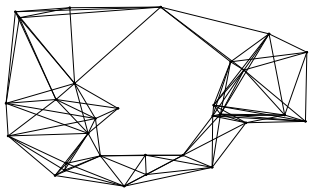
2.4.2 Non-triviality of the stationary distribution

In the previous section it has been shown that the finite Bak-Sneppen model converges weakly to a unique stationary distribution (dependent only on the graph). However, nothing has been said about the nature of this distribution. In general, it is difficult to say anything specific, but some general and asymptotic results are possible. Note that the stationary distribution of a Bak-Sneppen model on a complete graph is trivially uniform product measure, $U(0, 1)^N$. The following results comes from a comparison between Bak-Sneppen models on arbitrary graphs and complete graphs.

Proposition 2.24. *The stationary distribution* $\pi \geq_{st} U(0, 1)^N$.

Proof: This follows immediately from the coupling proof of Theorem 3.2 in Chapter 3. □

Proposition 2.25. *Each marginal has full support.*



Proof: Every vertex is a.s. minimal infinitely often. After it is minimal, the vertex's fitness is distributed $U(0, 1)$. □

Proposition 2.26. *If G is transitive, then π has identically distributed marginals.*

Proof: This follows almost immediately from the main result of the previous section. Assume that π doesn't have identically distributed marginals. Let $\pi^\star \neq \pi$ be a rotation of π. Since π is a stationary distribution, $\pi^\star$ must also be a stationary distribution. This is a contradiction, therefore π must have identically distributed marginals. □

However, the real interest is looking at the limit of π_N as $N \to \infty$. Considering again Figures 2.1 and 2.2, it appears that the fitnesses are (apart from near the minimal fitness) independent and uniformly distributed above some threshold.

Theorem 2.27. *Let G_N be a sequence of graphs such that $|G_N| \to \infty$ and $\Delta(G_N) \leq \Delta \ \forall N$, where $\Delta(G_N)$ is the maximal degree of G_N. Let π_N denote the stationary distribution of a Bak-Sneppen on G_N. Let π_N^i be the i-th marginal distribution function of π_N. Then $\lim_{N\to\infty} \pi_N^i(p) = 0$ $\forall p < \frac{1}{\Delta+1}$.*

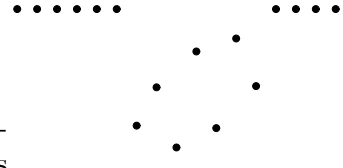
Proof: The approach is to consider the Bak-Sneppen model as a sequence of p-avalanches and uniformly bound, in N, the expected number of active vertices. This is done using the comparison with a branching process used in Section 2.3. Then as $N \to \infty$, the proportion of active vertices tends to zero, yielding the result.

Since all p-avalanches are finite on finite graphs, we wait until all fitnesses are greater than p. Alternatively, we can take our initial fitness distribution to be uniform $(p, 1)$. We can then decompose the model into a sequence of p-avalanches. We denote the duration of avalanche i by a_i and the total volume of avalanche i by A_i. Here the total volume of an avalanche is defined to be the number of active vertices at each time step summed over the duration of the avalanche. The average number of active vertices up to and including the n-th avalanche is

$$\frac{\sum_{i=1}^n A_i}{\sum_{i=1}^n a_i}.$$

This tends towards the expected number of active vertices, as $n \to \infty$, a.s. by the SLLN as the $\{A_i\}_{i\geq 1}$ are i.i.d. Since each avalanche has duration at least 1,

$$\frac{\sum_{i=1}^n A_i}{\sum_{i=1}^n a_i} \leq \frac{\sum_{i=1}^n A_i}{n} \to \mathbb{E}(A_1).$$



Thus $\mathbb{E}(A_1)$ gives an upper bound on the expected number of active vertices. This value varies for different G_N. However, the branching process

comparison shows that $\mathbb{E}(A_1)$ is finite whenever p is less than the critical value of a Galton-Watson branching process with binomial $(\Delta + 1, p)$ offspring distribution, see Corollary 2.15 or [43]. □

Theorem 2.27 uses the branching process lower bound to get a bound on the expected volume of an avalanche on all graphs G_N. To get the tighter bound of $p_c^r(G)$, where $G_N \to G$, one requires specific results about avalanches only shown for $\mathbb{Z}$, see Section 2.2. The following proposition comes from [38].

Proposition 2.28. *Suppose there exists $0 < p_c < 1$, such that for any $p < p_c$,*

$$\limsup_{N\to\infty} \limsup_{n\to\infty} G_N(n,p) = 0,$$

and for any $p > p_c$,

$$\limsup_{N\to\infty} \limsup_{n\to\infty} G_N(n,p) = 1.$$

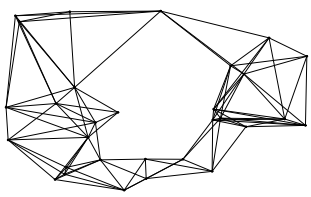
Then the limit distribution exists and is equal to the independent product of $U(p_c, 1)$ distributions.

For the special case of the Bak-Sneppen model on $\mathbb{Z}$, this has important consequences.

Theorem 2.29. *If $p_c^p(\mathbb{Z}) = p_c^r(\mathbb{Z})$, then the conditions of Proposition 2.28 are satisfied for $p_c = p_c^p(\mathbb{Z})$.*

2.4.3 Infinite Graphs

Let π_N be the stationary distribution of the Bak-Sneppen model on Λ_N. It is widely believed that π_N converges to product measure with uniform $(p_c^p, 1)$ marginals. This limit would be the natural candidate for the stationary distribution of the Bak-Sneppen model on $\mathbb{Z}$, especially when considering the limiting results from Section 2.2. However, this distribution is unreasonable for a Bak-Sneppen model on $\mathbb{Z}$. There is almost surely no minimal fitness and so the dynamics of the model are not defined.



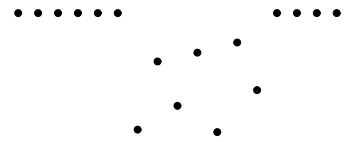
In this situation it is worth drawing an analogy between random walks. A random walk on a connected finite graph has a unique stationary distribution. However, on a infinite transitive graph, no such stationary distribution exists and the best one can do is find a stationary measure (e.g. every possible location has measure c for some $c > 0$).

I believe there is a similar situation for the Bak-Sneppen model in that the location of the minimal fitness is equally likely to be any vertex in the temporal limit and no stationary distribution exists.

2.5 Conclusions and Open Problems

Up to the present time, the mathematical research on the Bak-Sneppen model has been rather limited. The special case of models on finite circles Λ_N and $\mathbb{Z}$ is considerably easier to analyse than the BS model on other graphs. On this graph there is essential one main goal and open problem: the proof that $p_c^r(\mathbb{Z}) = p_c^p(\mathbb{Z})$. The proof of this is not only highly desirable for its own right, but also because of the consequences of such a proof. This result would rigorously demonstrate the qualitative shape of the limiting stationary distribution. This would form a vital first step in considering the claims of power law behaviour about the model.

When considering the Bak-Sneppen model away from $\mathbb{Z}$, it is natural to see if the qualitative behaviour of the model is different. In this report we have seen that it is indeed possible to dramatically change the behaviour by considering other graphs. However, certain results about the model that are not obtainable for $\mathbb{Z}$ maybe be achievable for $\mathbb{Z}^d$ for large d or on regular tree with a large number of offspring by making the comparison to the mean-field model rigorous.



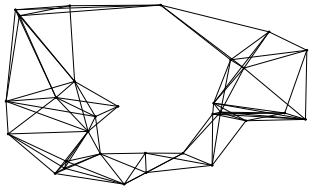
Chapter 3

Maximal avalanches in the Bak-Sneppen model

This chapter studies the durations of the avalanches in the maximal avalanche decomposition of finite Bak-Sneppen models. We show that all the avalanches in this maximal decomposition have infinite expectation, but only 'barely' in the sense that if we made the appropriate threshold a tiny bit smaller (in a certain sense), then the avalanches would have finite expectation. The first of these results is somewhat surprising, since simulations suggest finite expectations. This chapter has been published as a joint paper with Meester and van der Wal [42].

3.1 Introduction and main results

The Bak-Sneppen model was originally introduced as a simple model of evolution by Per Bak and Kim Sneppen [5]. Their model can be defined as follows. There are N species arranged on a circle with a random *fitness*, independent and uniformly distributed on $(0, 1)$, assigned to each species. At each discrete time step the system is updated by locating the lowest fitness and replacing this fitness and those of its two neighbours by independent and uniform $(0, 1)$ random variables. We think of the N species as the vertices of a circular graph Λ_N with N vertices.



This model can be defined on any finite connected graph, with the neighbours of each vertex being determined by the structure of the graph. For us, only the number of vertices of such a graph will be important (and the fact that it is connected), and we define $\mathcal{G}_N$ as the collection of con-

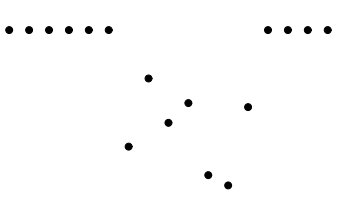
nected graphs with N vertices.

One of the ways to analyse the Bak-Sneppen model is to break it down into a series of *avalanches*. An avalanche at a *threshold* b (also called a *b-avalanche*) is said to occur between times s and $s+t$ if at time s all the fitnesses are equal to or greater than b with at most one vertex where equality holds, and time $s+t$ is the next time after s at which this occurs. In the literature a number of different types of avalanches has been proposed. Our definition is consistent with a number of other papers [5, 11, 37, 10], although b is not always used to denote the threshold in precisely the same way; other types of avalanches have for instance been defined in [24, 18]. Note that if we have a minimum fitness value of b, then we can choose any value up to (and including) b to be our avalanche threshold. Furthermore, it is the threshold (and not the exact initial values of the model) that determines the behaviour of an avalanche. Once we have used the initial fitnesses to find out the minimal fitness and its location, all other information can be discarded for the purposes of analysing individual avalanches.

One of the main reasons for interest in the Bak-Sneppen model is that it gives an example of *self-organised criticality* (SOC). Spatial probabilistic models (such as percolation) typically have a parameter that needs to be set before running the model, for example the probability of an edge being open in bond percolation. Often this parameter can be set to a particular value, the critical value, where unusual behaviour is observed. Models at criticality tend to exhibit scale-free or fractal behaviour, that is, the model looks 'the same' when viewed at different levels of magnification. Their behaviour often obeys power laws. However, this behaviour is very unstable, as the parameter has to be set exactly to the critical value.

Many observable phenomena seem to obey power laws, the relationship between the frequency of earthquakes and their sizes is perhaps the most commonly cited example. However, models that require a finely tuned parameter lack the robustness to explain such phenomena. This led Per Bak (among others) to consider models that tune themselves automatically to this critical state, coining the phrase self-organised criticality [4]. Interested readers should also have a look at Per Bak's book on the subject [1].

Returning to the original Bak-Sneppen model (on the circle), the notion of an avalanche helps to explain the criticality of the model. For large N there appears to be a threshold b_c, close to $2/3$, such that after a while, the dynamics appear to consist of consecutive avalanches at b_c, and in addition, these avalanches seem to exhibit power law behaviour in the sense that both duration and



range can be described by power laws [5]. The threshold b_c is not set beforehand; the model seems to organise *itself* into this state. See Figure 3.1 for a typical snapshot of the Bak-Sneppen model in stationarity, with $N = 300$. On the horizontal axis we have the 300 vertices, with the dots representing the fitnesses of the vertices.

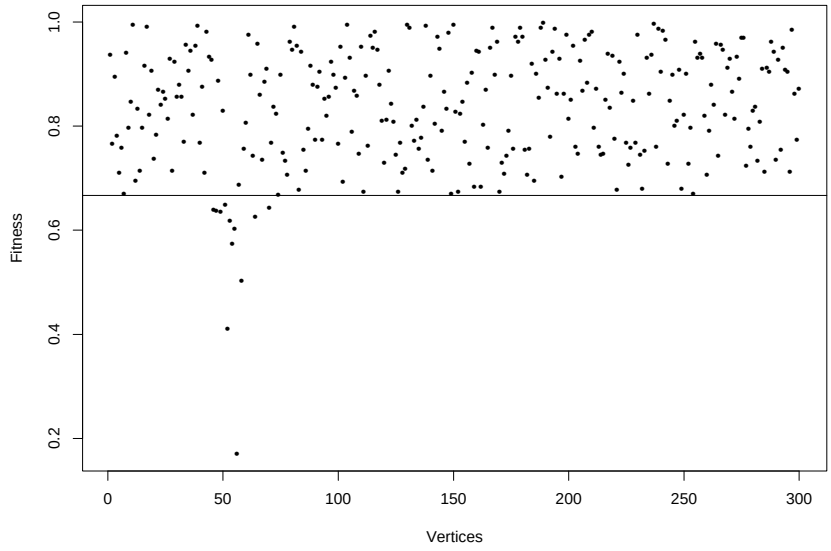
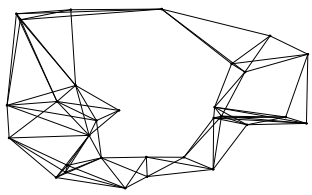


Figure 3.1: A snapshot of the fitnesses in the Bak-Sneppen model in stationarity.

The Bak-Sneppen model can be thought of as a sequence of consecutive avalanches. Since different avalanche thresholds can be chosen, there are a number of ways to perform such an *avalanche decomposition*. One common approach is to consider the model as a series of avalanches at some *fixed* threshold level b, see for instance [37].



There has also been considerable attention reserved for the so-called *maximal avalanche decomposition*. Here the first avalanche threshold is defined to be the minimum fitness value from the initial fitness values. After this and every subsequent avalanche, another avalanche begins with the threshold chosen to be the new minimal value of the model; this is the maximal threshold choice.

It is clear that this will lead to the Bak-Sneppen model being seen as a series of avalanches at strictly increasing thresholds. The *gap function* at time s, $G(s)$, is defined to be the avalanche threshold at time s [10]. The gap function is a stepwise increasing function which jumps to a new value each time an avalanche finishes. Note that for all finite systems the gap function tends to 1 almost surely. Figure 3.2 shows a realisation of the gap function represented by the line, with the dots being the minimum fitness values at each time step. The initial fitnesses were independent and uniform $(0, 1)$ distributed.

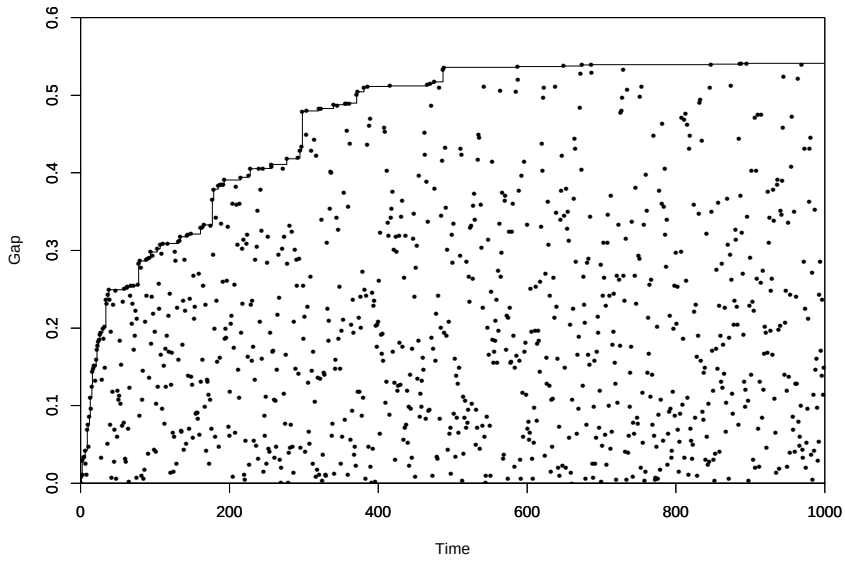


Figure 3.2: A realisation of the gap function when $N = 100$.

In this report we investigate the expected durations of avalanches in the maximal avalanche decomposition. One reason for looking at the maximal avalanche decomposition is to gain insight into how the Bak-Sneppen model tends towards criticality. An alternative approach to the same question is the so called *gap equation*. The gap equation is a conjectured differential equation that describes the behaviour of the gap function. It has been commonly referred to in the physics literature [10, 19] and is only informally defined. It refers to a different form of

the gap function and the gap equation is not examined in this paper.

Previous mathematical literature on the model concentrated on the expected duration of an avalanche at a fixed and non-random threshold b [38]. Their results include a number of useful monotonicity results, as well as an explicit differential equation relating the expected duration of avalanches to their expected range. In this paper, we study the avalanches at random thresholds which appear in, or are strongly related to, the thresholds in the maximal avalanche decomposition.

On Λ_N (or any transitive graph), the threshold is the only variable needed in order to determine the distribution of an avalanche's duration. By this we mean that the durations of two avalanches on a transitive graph are identically distributed if their thresholds are the same. On a non-transitive graph, the origin of the avalanche (the vertex which has the minimal fitness initially) also plays a role. However, one can still talk about the distribution of the initial avalanche with a random origin.

Consider the Bak-Sneppen model on $G_N \in \mathcal{G}_N$. Concentrating first on the *initial* avalanche in the maximal decomposition, we see that the initial threshold is the minimum of N independent uniform $(0, 1)$ random variables. To be more explicit, we have an avalanche with random threshold B whose density $h_N(b)$ is given by

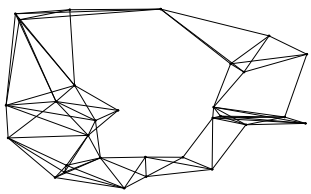
$$h_N(b) = N(1-b)^{N-1}, \qquad 0 \leq b \leq 1.$$

Letting $D(G_N)$ denote the random duration of the initial avalanche on G_N, we prove the following theorem.

Theorem 3.1. *The expected duration of the first avalanche on G_N is infinite, i.e. $\mathbb{E}(D(G_N)) = \infty$, for any $G_N \in \mathcal{G}_N$.*

One consequence of this result (for transitive graphs) is that any subsequent avalanche also has infinite expected duration, as its threshold is stochastically larger. Hence the gap function consists of a sequence of steps, each of which has infinite expected length.

The usual way to analyse the Bak-Sneppen model has been to run computer simulations. Compared to these simulations, our result seems somewhat surprising, since divergent behaviour is not typically noticeable under numerical simulations of the model, especially when N is large. This is because the long avalanches that are behind this result occur when the (random) threshold B is high, which is exponentially unlikely in N. If one were to run computer simulations of the initial avalanche in order to estimate its expected duration, it would



still be possible to detect this, but only from the dramatic variability of these estimations (even when a very large number of simulations are used). Theorem 3.1 is, therefore, an example of the value of analytic methods, as only very careful interpretation of computer simulations would lead one to suspect this result.

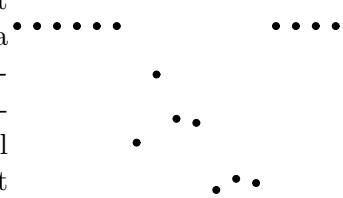
We decided to perturb the avalanche threshold by making it stochastically smaller and see whether this would lead to convergence. It turns out that $\mathbb{E}(D(G_N))$ is 'barely infinite' in that making the threshold a tiny bit stochastically smaller (where this reduction tends to 0 as N tends to ∞) yields finite expected durations. To be precise, we denote by $D_n(G_N)$ the expected duration of an avalanche at a threshold which is set by the minimum of n uniform $(0,1)$ random variables on G_N. In this notation, the previously used $\mathbb{E}(D(G_N))$ can now be written as $D_N(G_N)$, with Theorem 3.1 now stating that $D_N(G_N) = \infty$. We prove the following result.

Theorem 3.2. *For all $G_N \in \mathcal{G}_N$, an avalanche from a threshold chosen as the minimum of $n > N$ independent uniform $(0,1)$ random variables has finite expectation, i.e. $D_n(G_N) < \infty$ for all $n > N$.*

So just adding one uniform random variable when setting the threshold is enough to get a finite expected duration.

However, it is possible to show that under certain conditions all *further* avalanches have infinite expected duration. Recall that on G_N, setting the threshold as the minimum of N independent uniform (0,1) random variables, gives infinite expected duration. If all the fitnesses (except the minimum) are independent and uniformly distributed above the threshold at the start of the avalanche, then at the end of the avalanche all the vertices will again be independent and uniformly distributed above the threshold. So even if you fix b and choose your fitnesses to be uniform above it, it follows from Theorem 3.1 that the next avalanche will have infinite expected duration.

A more general, but weaker form of this result applies when we drop the condition that the fitnesses had to be nicely distributed at the start of the avalanche. All the vertices updated by the avalanche will be independent and uniformly distributed above the threshold at the end of the avalanche. So once we have had a spanning avalanche (one that updates every vertex in the system during its duration) all subsequent avalanches (from maximal thresholds) will have infinite expected duration, no matter what initial fitness values are taken.



3.2 Proof of Theorem 3.1

We prove that the initial maximal avalanche has infinite expected duration. This is done by first proving the result for the special case of $G_N = \Lambda_N$ and then explaining how to modify the proof to generalise the result onto all graphs $G_N \in \mathcal{G}_N$.

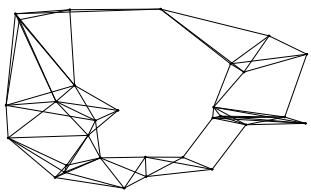
It is essentially a trivial result that $D_3(\Lambda_3) = \infty$. Indeed, at each time step the fitnesses are independently and identically distributed and so this is just a record values result [30]. Alternatively you can find an explicit expression for the duration of an avalanche at a fixed threshold and then integrate out over the distribution of B, the random avalanche threshold. For $N > 3$ a more subtle approach is required since we have neither independence nor identically distributed minimal fitnesses. In the proof for general N, we use the observation that if at any time all the fitnesses are below the (random) threshold B, then the avalanche cannot stop before at least N update values greater than B have been observed. We calculate the expected time taken for N such updates to be observed between times when all fitnesses are below the threshold.

Before we start on the main body of the proof we need some definitions. Let $f_i(n)$ denote the fitness of vertex i at time step n, where the initial fitnesses, corresponding to $n = 0$, are independent and uniform on $(0, 1)$. We generate our updates from a sequences of independent uniform $(0, 1)$ random variables, $U_k, k = 0, 1, 2, \ldots$ as follows: if vertex i has the minimal fitness at time n, then the fitnesses at positions $(i-1, i, i+1)$ are replaced by $(U_{3n}, U_{3n+1}, U_{3n+2})$, giving the fitnesses at time $n + 1$.

We divide the avalanche into time *blocks* of length $2N$, i.e. block m contains the time steps starting with $2mN$ and finishing with $2(m+1)N-1$, for $m = 0, 1, 2, \ldots$ We call a block *bad* if all the updates in the block are below the threshold B and are in decreasing order. So, block m is bad if

$$B > U_{6mN} > U_{6mN+1} > U_{6mN+2} > \cdots > U_{6(m+1)N-1}.$$

We let s_r denote the (random) time just after the r^{th} bad block, for $r = 1, 2, \ldots$ The next lemma tells us that after a bad block, all fitnesses are below B.



Lemma 3.3. *We have that $f_i(s_r) < B$ for all $0 \leq i \leq N-1$ and $r = 1, 2, \ldots$*

Proof: Consider the state of the system at the beginning of the r^{th} bad block. We update the vertex with the minimal fitness, together with its

two neighbours. Since the block is bad, the three updated vertices now have fitness below the threshold. We now have two possible cases, either one of the new fitnesses is the minimum, or another vertex has the minimal fitness. If the minimum is drawn from a newly updated vertex then the avalanche will continue (for the duration of the bad block) round the circle of vertices, since after another update the minimum among the three updated vertices will be lower than the minimum at the previous step and so will in fact be the global minimum. The second case can not occur more than N times before the first case occurs. Once we have the minimum among new fitnesses, it can not take more than N vertices for the first case to cover the entire vertex set. Hence after $2N$ updates all the vertices will have fitnesses generated by new updates. Thus all fitnesses will be below the threshold, as required. □

We call a period between the endpoints of two bad blocks an *epoch*; hence the m^{th} epoch corresponds to the time interval $[s_m, s_{m+1} - 1]$, for $m = 1, 2, \ldots$ We call an epoch *good* if among the updating random variables corresponding to that epoch there are at least N updating random variables greater than B. So, the m^{th} epoch is good if

$$\sum_{k=3s_m}^{3s_{m+1}-1} \mathbb{I}\{U_k \geq B\} \geq N,$$

where $\mathbb{I}$ denotes the indicator function. We let S be the index of the first good epoch. Now observe that *given* that block 0 is bad, the first avalanche can only finish during a good epoch. This indicates that the expectation of S plays an important role in our analysis.

Lemma 3.4. $\mathbb{E}(S) = \infty$.

Proof: The key to the proof is to compute the expectation of S given that $B = b$. Indeed, conditioned on $B = b$, S has a geometric distribution with a successful trial being a good epoch. This is because conditional on the threshold, the lengths of the epochs are independent and identically distributed. We let $p(b)$ be the probability that the first epoch (or any other, for that matter) is good, given that $B = b$. Integrating over the range of b, we then have that

$$\mathbb{E}(S) = \int_0^1 \frac{N(1-b)^{N-1}}{p(b)} db,$$

since B is distributed as the minimum of N independent uniform $(0, 1)$ random variables. It remains to estimate $p(b)$. Since we want to bound

the integral from below, we have to bound $p(b)$ from above. We let F_n be the event that the first epoch does not finish before or at time $s_1 + n$, and write G_n for the event that

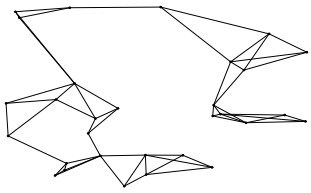
$$\min\{\ell; \sum_{k=3s_1}^{\ell} \mathbb{I}\{U_k \geq B\} \geq N\} = n,$$

that is, G_n is the event that the n^{th} update random variable since the start of the first epoch - ordered naturally as $U_{3s_1}, U_{3s_1+1}, \ldots$ - is the N^{th} above b. We can now write:

$$p(b) = \sum_{n=N}^{\infty} \mathbb{P}(G_n | B = b)\mathbb{P}(F_{\lfloor n/3 \rfloor} | G_n, B = b).$$

Now observe that $\mathbb{P}(G_n|B = b) = \binom{n-1}{N-1}(1-b)^N b^{n-N}$ which we bound from above by $\binom{n-1}{N-1}(1-b)^N$. Furthermore, conditioning on $B = b$ and G_n means that we randomly select $N-1$ updates among the $n-1$ first updates after s_1 which are declared to be the updates above b; the only information about all the other updates is that they are below b. No matter the precise realisation, this implies that there are at least (for $n > 6N^2$) $\lfloor \frac{n-1}{6N} \rfloor - (N-1)$ blocks so far for which our only information is that they have all their corresponding updates below b. Hence the (conditional) distribution of the updates in these blocks are i.i.d. uniform on $(0, b)$, and any of these blocks is bad with a (conditional) probability which is uniformly bounded below by a constant $c_1(N)$, say. This then gives:

$$\begin{aligned} p(b) &\leq \sum_{n=N}^{6N^2} \binom{n-1}{N-1}(1-b)^N + \\ &\quad \sum_{n=6N^2+1}^{\infty} \binom{n-1}{N-1}(1-b)^N(1-c_1(N))^{\lfloor \frac{n-1}{6N} \rfloor - (N-1)} \\ &\leq (1-b)^N \left(\sum_{n=N}^{6N^2} \binom{n-1}{N-1} + \sum_{n=6N^2+1}^{\infty} n^N (1-c_1(N))^{\lfloor \frac{n-1}{6N} \rfloor - (N-1)} \right) \\ &= c(N)(1-b)^N, \end{aligned}$$



for a suitable constant $c(N) < \infty$.

We can now substitute our upper bound for $p(b)$ back into our integral for $\mathbb{E}(S)$, giving

$$\begin{aligned}\mathbb{E}(S) &= \int_0^1 \frac{N(1-b)^{N-1}}{p(b)}db \\ &\geq \frac{N}{c(N)}\int_0^1 \frac{(1-b)^{N-1}}{(1-b)^N}db \\ &= \frac{N}{c(N)}\int_0^1 \frac{1}{1-b}db \\ &= \infty.\end{aligned}$$

□

Let A be the event that block 0 is bad. We already noticed that if block 0 is bad, the first avalanche can only end during a good epoch. This gives

$$\mathbb{E}(D(\Lambda_N)) \geq \mathbb{E}((D(\Lambda_N)|A)\mathbb{P}(A) \geq \mathbb{E}(S|A)\mathbb{P}(A),$$

and since $\mathbb{P}(A) > 0$ it suffices to prove that $\mathbb{E}(S|A) = \infty$.

Lemma 3.5. *We have that* $\mathbb{E}(S|A) = \infty$.

Proof: Writing $F_{B|A}$ for the conditional distribution function of B given the event A, we have that

$$\mathbb{E}(S|A) = \int_0^1 \mathbb{E}(S|A, B = b)dF_{B|A}(b). \tag{3.1}$$

We claim now that

$$\mathbb{E}(S|A, B = b) = \mathbb{E}(S|B = b). \tag{3.2}$$

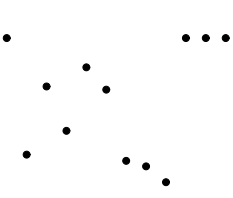
To see this, observe that when we condition on $B = b$, further knowledge about updates in block 0 is irrelevant for the distribution of S; indeed, the distribution of S depends only on B, with $\mathbb{E}(S|B = b)$ non-decreasing in b.

Furthermore, we claim that

$$F_{B|A}(b) \leq F_B(b), \tag{3.3}$$

where F_B denotes the distribution function of B. To see this, we write $U = \max\{U_0, \ldots, U_{6N-1}\}$ and argue as follows:

$$\begin{aligned}F_{B|A}(b) &= \mathbb{P}(B \leq b|A) \\ &= \mathbb{P}(B \leq b|B > U_0 > U_1 > \cdots > U_{6N-1}) \\ &= \mathbb{P}(B \leq b|B > U, U_0 > U_1 > \cdots > U_{6N-1}).\end{aligned}$$



Since the event $\{U_0 > U_1 > \cdots > U_{6N-1}\}$ is independent of (U, B), we can remove it from the conditioning event, and we find that

$$\begin{aligned} F_{B|A}(b) &= \mathbb{P}(B \leq b | B > U) \\ &\leq \mathbb{P}(B \leq b), \end{aligned}$$

since B and U are independent, and it is well known in general that conditioning a random variable B to be larger than another, independent, random variable U makes B stochastically larger.

Taking together (3.1), (3.2) and (3.3), with the fact that $\mathbb{E}(S|B = b)$ is non-decreasing in b, we finally obtain

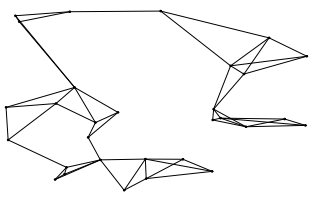
$$\mathbb{E}(S|A) \geq \int_0^1 \mathbb{E}(S|B = b) dF_B(b) = \mathbb{E}(S),$$

which by Lemma 3.4 is equal to infinity, proving the result. □

The proof above only gives the result for Λ_N. However, the only part of the proof that really utilised the circular graph structure was Lemma 3.3. If, for any graph G_N, an equivalent result to Lemma 3.3 can be proven, then the rest of the proof will follow as above (up to a few changes of constants). The goal is, therefore, to come up with a similar notion of a bad block that again ends with all fitnesses below the threshold. Below is a recipe for doing this for any graph G_N.

A *closed walk* in a graph is defined to be an ordered collection of vertices $(v_1, v_2, \ldots, v_n)$ such that v_i and v_{i+1} are neighbours for all $1 \leq i \leq n-1$, and v_n and v_1 are neighbours. Note that it is allowed to visit the same vertex or traverse the same edge on more than one occasion. For any finite connected graph it is possible to find a closed walk that visits every vertex of the graph at least once. Take a minimal such closed walk on G_N.

This walk is now used to update the Bak-Sneppen model. Recall that the updates are drawn from a sequence of independent uniform $(0, 1)$ random variables $U_0, U_1, U_2, \ldots$. Let the vertex with the initial minimal fitness be v. Locate v in the walk. It may occur more than once, in this case just pick one of these. Now update this vertex and its neighbours in such a way that the last random variable from the sequence is used to update the neighbour of v that occurs next in the walk. For example, if we use U_0, U_1 and U_2, then U_2 will be the new fitness of the vertex that comes next in the walk. Now we continue along the walk until we get to the vertex with the new minimal fitness and update as above.



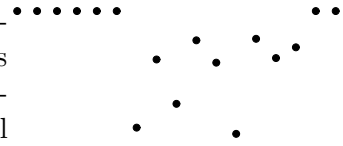
Now consider what happens if we have a bad block. Recall that a bad block is a block in our sequence of random variables where the random variables are decreasing and below the threshold. If, at any point during the bad block, the minimal fitness comes from the last random variable of the bad block, then the location of the minimal fitness will follow the walk updating all the vertices.

For any finite connected graph G_N, one can bound from above the number of random variables needed in a bad block to ensure that the minimal fitness comes from the last random variables of the bad block. Note that here we are measuring the length of a bad block in terms of random variables rather than update steps (as before). Denote this bound by l_1 (in the case of $G_N = \Lambda_N$, we used $l_1 = 3N$ in the proof of Lemma 3.3). Let l_2 be the number of random variables used to perform our walk (in the case of $G_N = \Lambda_N$, we used $l_2 = 3N$ in the proof of Lemma 3.3). Then a bad block of length $l = l_1 + l_2$ will ensure that all fitnesses are below the threshold as desired.

3.3 Proof of Theorem 3.2

Let $G_N \in \mathcal{G}_N$. Recall that Theorem 3.2 stated that an avalanche on G_N from a threshold distributed as the minimum of $N+1$ independent uniform random variables has finite expected duration. Our approach is to couple the Bak-Sneppen model with a similar and simpler model whose avalanche durations are more easily calculated. In the Bak-Sneppen model we have the vertices arranged on a circle and so each vertex has only two neighbours. For the Bak-Sneppen model on the *complete graph* K_N, updating the minimal vertex and its neighbours will update the entire graph. This yields a fairly uninteresting model, with no time dependence or self-organised critical behaviour in the infinite limit. However, the expected duration of an avalanche on this model from a threshold b is easy to calculate and is equal to $(1-b)^{-N} < \infty$.

Intuitively one might think that an avalanche on K_N from a threshold b would be longer than an avalanche on G_N from the same threshold, b. This is because the original model eliminates the smallest value and its neighbours, leaving some old fitnesses that are certainly greater than the minimum. It seems plausible that the distributions of these other vertices are stochastically larger than uniform $(0,1)$ random variables. This intuition is correct, as we will now demonstrate.



Since, as in the Bak-Sneppen model on Λ_3, we

have an explicit expression for the expected duration of an avalanche at a fixed threshold on K_N, we can calculate $D_n(K_N)$ by integration.

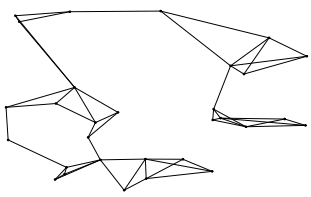
$$D_n(K_N) = \int_0^1 \frac{n(1-b)^{n-1}db}{(1-b)^N} < \infty, \text{ for all } n > N,$$

where the expression $n(1-b)^{n-1}$ is the density of the threshold which is set by the minimum of n uniform random variables. So we need to provide a coupling of the Bak-Sneppen models on K_N and G_N to show that the expected duration of a b-avalanche is shorter for G_N and therefore that the initial avalanche has finite expectation when $n > N$.

The main difficulty to overcome when coupling these two models together is that the time dependencies are completely different. The Bak-Sneppen model on G_N depends heavily upon the values at the previous step, whereas on K_N the fitnesses are completely independent of previous fitness values. Our approach is as follows. We consider an alternative way of constructing a Bak-Sneppen avalanche that does not fix the fitness values, but has the same duration distribution. We then demonstrate how this construction can be realised using a sequence of independent uniform $(0, 1)$ random variables and in this way coupled with the model on K_N. Readers should note that a similar approach to this has also been used to compare Bak-Sneppen avalanches to site percolation [43]. An explicit example of this coupling is given before proving that this coupling has the desired properties.

Consider a Bak-Sneppen b-avalanche on G_N. Pick a vertex uniformly at random to be the origin of the avalanche and give it fitness b. Let all the other fitnesses be uniformly distributed on $(b, 1)$. At the first update step we generate new fitnesses that are associated with (but not assigned to!) the origin and its neighbours. Furthermore, we generate new fitnesses for the remaining vertices, but this time they are not uniform $(0, 1)$, but uniform $(b, 1)$. We find the minimum of these values and assign it to the respective vertex.

Now, in contrast to the usual treatment, we discard the remaining fitness values and merely record the conditional distributions of the fitnesses given that the fitnesses must be greater than the given minimal value. We iterate this process by again generating new fitnesses corresponding to the minimal vertex and its neighbours and also generate new fitness values corresponding to the other vertices contained in G_N according to their conditional distributions. Again we locate and fix the minimal value before discarding the other fitnesses and recording their



conditional distributions (i.e given their previous distribution and that they are greater than the new minimal value). Note here that at each time step only one vertex has a fixed value, namely the minimal one.

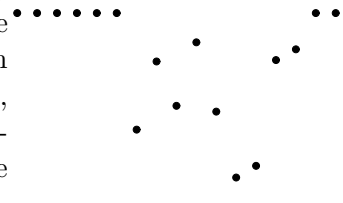
If at any point one wishes to have exact fitnesses for the Bak-Sneppen model, the associated fitness values can be fixed to all vertices. This is similar to the inductive step above, except that after locating and recording the minimum value the remaining fitnesses are fixed rather than discarded. Fixing all the fitness values ends this construction, since it requires the minimum fitness to be the only fixed value. So to continue the avalanche after the fixing, one must return to the original method for realising the Bak-Sneppen model.

Proposition 3.6. *The above construction of a b-avalanche has the same range and duration distributions as a Bak-Sneppen b-avalanche.*

Proof: The construction above follows the rules for a Bak-Sneppen model in that at each time step the vertex with the minimal fitness is updated along with the fitnesses of its neighbours. The only difference is the use of information. Traditionally the fitnesses of all the vertices are known at all time steps, but here we only keep track of distributions. By construction, these distributions are the same as the fitness distributions in a Bak-Sneppen avalanche if the precise fitness values were unknown. In particular, the location and magnitude of the minimal vertex has the correct distribution. Since the location and magnitude is at all times distributed the same as in a normal Bak-Sneppen avalanche, the required equivalence is now clear. The duration is determined by the magnitudes of the sequence of minima and the range by their respective locations. □

We remark that the idea of constructing the Bak-Sneppen process in this way is closely related to the 'locking thresholds representation' introduced in [37]. With this new construction of a b-avalanche we are now ready to describe the coupling.

We start by generating a sequence of uniform $(0,1)$ random variables, $X_1, X_2, \ldots$ that shall be used to update the fitnesses on K_N. Contrary to earlier notation, for this proof we enumerate the vertices by $1, 2, ..., N$; the fitnesses on K_N will be $X_1, X_2, ..., X_N$ respectively after the first time step, and $X_{N+1}, X_{N+2}, ..., X_{2N}$ after the second time step and so on. We use the same random variables to generate the fitnesses on G_N. However, in the Bak-Sneppen model on G_N, we do not always wish to use uniform $(0,1)$ random variables. Before they can be used, we have to transform our uniform $(0,1)$ random variables.

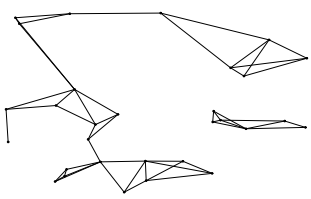


This is simply done by taking the inverse of the distribution functions (sometimes referred to as probability integral transformation).

It is perhaps worth describing the exact nature of these transformations more precisely. If you wish to generate a random variable with any invertible distribution function, F, then this can be done by taking a uniform $(0,1)$ distributed random variable, X, and taking $F^{-1}(X)$. So our approach here is to use this sort of transformation so that the same sequence (X_i) can be used for both the model on K_N and G_N.

We consider two cases. The first possibility is that the avalanche on K_N stops. In this case we fix the fitness values on G_N to check if the Bak-Sneppen avalanche has also stopped. This is done as above: the random variables are transformed to the correct distributions and then set to be the fitnesses of the vertices. If, on the other hand, the avalanche on K_N is still in progress, we carry on with the Bak-Sneppen avalanche on G_N according to our new construction.

We are now ready to give a specific example (in this case the Bak-Sneppen model on $G_N = \Lambda_N$), as we know how to update the fitness distributions in the coupled Bak-Sneppen model. Each time step is split into two lines in order to illustrate the mechanism of the coupling more clearly. In the first line we update on K_N and write down the marginal distributions for the Bak-Sneppen model at that update. On the next line the fitness values for K_N are used to calculate the exact minimum for the Bak-Sneppen model and to update the remaining marginal distributions. For brevity we use the shorthand $> b$ for $U(b,1)$ and > 5 for $U(X_5,1)$. We use an underscore to denote the minimum fitness value. It is important to remember that the minimum on K_N doesn't play an explicit role in determining the behaviour of the Bak-Sneppen model on G_N. The underscore is used here only to highlight the types of behaviour that you can observe from this coupling.



Since vertex 2 has the minimum value initially, the new fitnesses of vertices $1, 2$ and 3 will be uniform $(0,1)$. We also update the fitnesses on K_N. Note that in this example, at all times the minimal fitness on K_N is below b, so the avalanche doesn't stop during the steps we have detailed. So we don't fix the exact fitness values in the coupled

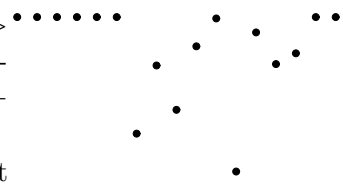
Bak Sneppen Model on K_5					Comments	Bak-Sneppen model on Λ_5				
1	2	3	4	5	Vertices	1	2	3	4	5
$> b$	$\underline{b}$	$> b$	$> b$	$> b$	Initial setup	$> b$	$\underline{b}$	$> b$	$> b$	$> b$
X_1	X_2	X_3	X_4	X_5	Update K_5	> 0	> 0	> 0	$> b$	$> b$
X_1	X_2	$\underline{X_3}$	X_4	X_5	Update Λ_5	> 3	> 3	$\underline{X_3}$	$> b$	$> b$
X_6	X_7	X_8	X_9	X_{10}	Update K_5	> 3	> 0	> 0	> 0	$> b$
$\underline{X_6}$	X_7	X_8	X_9	X_{10}	Update Λ_5	> 9	> 9	> 9	$\underline{X_9}$	$> b$
X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	Update K_5	> 9	> 9	> 0	> 0	> 0
X_{11}	$\underline{X_{12}}$	X_{13}	X_{14}	X_{15}	Update Λ_5	$> x$	$\underline{x}$	$> x$	$> x$	$> x$

Figure 3.3: Coupling with a Bak-Sneppen model on the complete graph

Bak-Sneppen model, but instead use the random variables, $X_1, X_2, ..., X_5$, to find the location of the minimal fitness in the Bak-Sneppen model. On K_N, X_3 takes the minimum value and so vertex 3 in the Bak-Sneppen model is minimal with fitness X_3. This is because out of the transformed values X_3 is necessarily the smallest, as the transformations to X_4 and X_5 make them larger and no transformations are required for X_1, X_2 and X_3. Since the avalanche is still in progress, $X_3 < b$, and so the distributions of the fitnesses of vertices 4 and 5 remain the same, whereas for vertices 1 and 2 they are now uniform $(X_3, 1)$.

The next time step follows on in a similar vein, except that after the transformations we find that X_9 is the minimal transformed value even though before transformation X_6 was smaller. This means that the location of the minimal vertex on the two models is now different. It is also worth noticing that X_9 must be greater than X_3 by the way that the fitness distributions have been updated. The final time step detailed gives an example of the more complicated behaviour that can be observed. Here vertex 2 is minimal in both models, but in the Bak-Sneppen model the fitness value at vertex 2 is x rather than X_{12}, where $x = X_9 + (1 - X_9)X_{12}$. This is because we have had to transform the value of X_{12} so that it had the correct distribution.

To conclude this example let us consider the case when the avalanche in the new model stops. This means that $\min(X_{11}, X_{12}, X_{13}, X_{14}, X_{15}) > b$. We would then fix the fitnesses for the Bak-Sneppen model giving $X_9 + (1 - X_9)X_{11}, X_9 + (1 - X_9)X_{12}, X_{13}, X_{14}, X_{15}$.



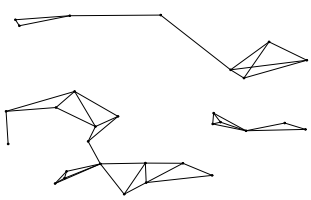
Having given a description of the coupling, it remains to confirm that this coupling has the de-

sired property: that the Bak-Sneppen avalanche on G_N can not outlive the avalanche on K_N.

Proposition 3.7. *The Bak-Sneppen avalanche finishes no later than the end of the new process.*

Proof: This proposition is proved by showing that the conditional distributions of the Bak-Sneppen model always belong to a certain family of distributions. Observe that initially we have one fixed fitness with all the others uniformly distributed above the threshold b. When a vertex or one of its neighbours is minimal its fitness distribution is reset to be uniform $(0, 1)$. The only other way the fitness distribution can be altered is by conditioning it on being bigger than a given value. Conditioning a uniform random variable to be bigger than a given value merely restricts the variable to be uniform on the part of its original range that is above the given value. It is obvious that a uniform $(0, 1)$ random variable is stochastically dominated by a uniform $(x, 1)$ random variable when $x \geq 0$. Thus, transforming the fitnesses from K_N to make them applicable to the Bak-Sneppen model on G_N can never make them smaller. Hence, if the fitnesses on K_N are all greater than the threshold b, then so must the fitnesses of the Bak-Sneppen model on G_N. □

This coupling shows that $D_n(K_N) < \infty$ implies $D_n(G_N) < \infty$ and hence we have proved Theorem 3.2. □



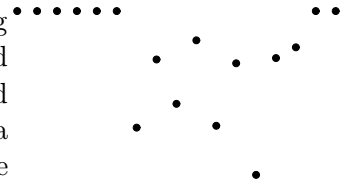
Chapter 4

Bounds for avalanche critical values of the Bak-Sneppen model

This chapter studies the Bak-Sneppen model on locally finite transitive graphs G, in particular on $\mathbb{Z}^d$ and on T_Δ, the regular tree with common degree Δ. We show that the avalanches of the Bak-Sneppen model dominate independent site percolation, in a sense to be made precise. Since avalanches of the Bak-Sneppen model are dominated by a simple branching process, this yields upper and lower bounds for the so-called avalanche critical value $p_c^{BS}(G)$. Our main results imply that $\frac{1}{\Delta+1} \leq p_c^{BS}(T_\Delta) \leq \frac{1}{\Delta-1}$, and that $\frac{1}{2d+1} \leq p_c^{BS}(\mathbb{Z}^d) \leq \frac{1}{2d} + \frac{1}{(2d)^2} + O\big(d^{-3}\big)$, as $d \to \infty$. This chapter has been published as a joint paper with Meester and Nuyens [43].

4.1 Introduction and main results

The Bak-Sneppen model was originally introduced as a simple model of evolution by Per Bak and Kim Sneppen [5]. Their original model can be defined as follows. There are N species (vertices) arranged on a circle, each of which has been assigned a random *fitness*. The fitnesses are independent and uniformly distributed on $(0, 1)$. At each discrete time step the system evolves by locating the lowest fitness and replacing this fitness, and those of its two neighbours, by independent and uniform $(0, 1)$ random variables. We say that a vertex whose fitness is changed by this procedure has been *updated*.

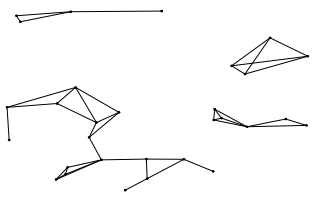


It is not particularly significant that the underlying graph of the model is the circle, or $\mathbb{Z}$ in the thermodynamic limit. Bak-Sneppen models can be defined on a wide range of graphs using the same update rule as above: the vertex with minimal fitness and its neighbours are updated. Unlike particle systems such as percolation or the contact process, the Bak-Sneppen model has no tuning parameter. Therefore, it has been described as exhibiting self-organised critical behaviour, see [41] for a discussion.

One of the ways to analyse Bak-Sneppen models is to break them down into a series of *avalanches*. An avalanche from a threshold p, referred to as a p-avalanche, is said to occur between times s and $s+t$ if at time s all the fitnesses are equal to or greater than p with at most one vertex where equality holds, and time $s+t$ is the first time after s at which all fitnesses are larger than p. The vertex with minimal fitness at time s is called the *origin* of the avalanche. A p-avalanche can be considered as a stochastic process in its own right. The key feature of the origin is that it has the minimal fitness (as it will be updated immediately). Hence, we can consider its fitness to be any value, as long as this value is minimal. Vertices with fitness below the threshold are called *active*, others are called *inactive*. Note that the exact fitness value of an inactive vertex is irrelevant for the avalanche, since this value can never be minimal during the avalanche. This motivates the following formal definition of an avalanche.

Definition 4.1. *A p-avalanche with origin v on a graph G (with vertex set $V(G)$) is a stochastic process with state space $\{[0,p]^A, A \subset V(G)\}$ and initial state $p^{\{v\}}$. The process follows the update rules of the Bak-Sneppen model. Any vertex with a fitness smaller than or equal to p is included. Any vertex with a fitness larger than p is not included. The process terminates when it is the empty set.*

Studying avalanches has considerable advantages. A Bak-Sneppen model on an *infinite* graph is not well-defined: when there are infinitely many vertices, there may not be a vertex with minimal fitness. However, Bak-Sneppen *avalanches* can be defined on any locally finite graph as follows: at time 0 all vertices have fitness 1, apart from one vertex, the origin of the avalanche, which has fitness p. We then apply the update rules of the Bak-Sneppen model, until all fitnesses are above p. This is consistent with our previous notion, as it is only the fitnesses updated during the avalanche that determine the avalanche's behaviour. The ability to look directly at infinite graphs is very desirable, because the most interesting behaviour of the Bak-Sneppen model is observed in the limit



as the number of vertices in the graph tends to infinity.

In the literature alternative types of avalanches have been proposed, see [24, 18]. The definition given here corresponds to the most commonly used notion of an avalanche and was introduced by Bak and Sneppen [5]. For a more thorough coverage readers are directed to Meester and Znamenski [37, 38]. Note that unlike the Bak-Sneppen model itself, the avalanches do have a tuning parameter, namely the threshold p.

In this paper, we look mainly at *transitive* graphs. The behaviour of an avalanche on a transitive graph is independent of its origin: an avalanche with origin at vertex v behaves the same as an avalanche with origin 0. When analysing avalanches on transitive graphs, it is therefore natural to talk about a typical p-avalanche without specifying its origin. To analyse avalanches, some definitions are needed. The set of vertices updated by an avalanche is referred to as its range set, with the *range* being the cardinality of this range set. Letting $r_G^{BS}(p)$ denote the range of a p-avalanche on a transitive graph G, we define the *(avalanche) critical value* of the Bak-Sneppen model as

$$p_c^{BS}(G) = \inf\{p : \mathbb{P}(r_G^{BS}(p) = \infty) > 0\}. \tag{4.1}$$

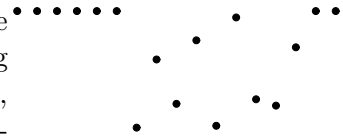
Numerical simulations [5] suggest that the stationary marginal fitness distributions for the Bak-Sneppen model on N sites tend to a uniform distribution on $(p_c^{BS}(\mathbb{Z}), 1)$, as $N \to \infty$. It has been proved in [38] that this is indeed the case if $p_c^{BS}(\mathbb{Z}) = \widehat{p}_c^{BS}(\mathbb{Z})$, where $\widehat{p}_c^{BS}(\mathbb{Z})$ is another critical value, based on the expected range, and is defined as

$$\widehat{p}_c^{BS}(G) = \inf\{p : \mathbb{E}[r_G^{BS}(p)] = \infty\}. \tag{4.2}$$

It is widely believed, but unproven, that these two critical values are equal.

It should now be clear that knowledge about the value of $p_c^{BS}(G)$ is vital in determining the self-organised limiting behaviour of the Bak-Sneppen model, even though there is no tuning parameter in the model. Although in this paper we focus on the critical value (4.1), our bounds for the critical value (4.1) also hold for the critical value (4.2), see Section 4.6.

The approach of this paper is to compare Bak-Sneppen avalanches with two well-studied processes, namely branching processes and independent site percolation. A simple comparison with branching processes gives a lower bound on the critical value, whereas a more complex comparison with site percolation gives an upper bound. To warm up, we first give the (easy) lower bound.



Proposition 4.2. *On any locally finite transitive graph G with common vertex degree Δ, we have*

$$p_c^{BS}(G) \geq \frac{1}{\Delta + 1}.$$

Proof: At every discrete time step of the system, we draw $\Delta + 1$ independent uniform $(0, 1)$ random variables to get the new fitnesses of the vertex with minimal fitness, and of its Δ neighbours. Each of these $\Delta + 1$ new fitnesses is below the threshold p with probability p, independent of each other. This induces a coupling with a simple branching process with binomial $(\Delta + 1, p)$ offspring distribution, where every active vertex in the Bak-Sneppen avalanche is represented by at least one particle in the branching process. Hence, if the branching process dies out, then so does the Bak-Sneppen avalanche. Therefore the critical value of the Bak-Sneppen avalanche can be no smaller than the critical value of the branching process. □

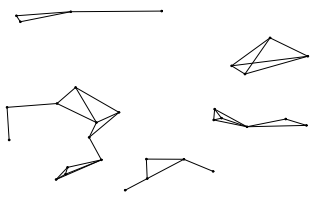
The main result of this paper is the following upper bound for the critical value $p_c^{BS}(G)$ of the Bak-Sneppen model on a locally finite transitive graph G. The critical value for independent site percolation on G is denoted by $p_c^{site}(G)$. We recall that for site percolation on G with parameter p, the probability of an infinite cluster at the origin is positive for all $p > p_c^{site}(G)$, and 0 for all $p < p_c^{site}(G)$.

Theorem 4.3. *On any locally finite transitive graph G, we have*

$$p_c^{BS}(G) \leq p_c^{site}(G).$$

This result implies that on many locally finite transitive graphs, p_c^{BS} is non-trivial. For the Bak-Sneppen avalanche on $\mathbb{Z}$, Theorem 4.3 gives a trivial upper bound, but in this case we know from [37] that $p_c^{BS}(\mathbb{Z}) \leq 1 - \exp(-68)$.

Since the critical value of site percolation on T_Δ, the regular tree with common degree Δ, equals $1/(\Delta - 1)$, the following corollary holds.



Corollary 4.4. *The critical value of the Bak-Sneppen model on a regular tree T_Δ, with common degree Δ, satisfies*

$$\frac{1}{\Delta + 1} \leq p_c^{BS}(T_\Delta) \leq \frac{1}{\Delta - 1}.$$

Applying the expansion for the critical value of site percolation on $\mathbb{Z}^d$ given by Hara and Slade [35], we also have the following corollary.

Corollary 4.5. *The critical value of the Bak-Sneppen model on $\mathbb{Z}^d$ satisfies*

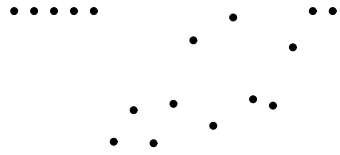
$$\frac{1}{2d+1} \leq p_c^{BS}(\mathbb{Z}^d) \leq \frac{1}{2d} + \frac{1}{(2d)^2} + O\big(d^{-3}\big), \qquad d \to \infty.$$

The paper is organised as follows. In Section 4.2 we take some preliminary steps by describing an alternative way of constructing a Bak-Sneppen avalanche. Section 4.3 uses this construction to couple the Bak-Sneppen avalanche and another stochastic process. The proof that the critical value of the Bak-Sneppen avalanche is larger than that of the coupled stochastic process is given in Section 4.4. The proof of Theorem 4.3 is completed in Section 4.5 where we show that the coupled process in fact constructs the cluster at the origin of site percolation with the origin always open. In Section 4.6, we discuss some implications and generalisations of our methods and results.

4.2 An alternative construction of the Bak-Sneppen model

In the introduction the Bak-Sneppen model was defined in its original format and then generalised to locally finite graphs. However, for our purposes it is more convenient to work with an alternative construction of the Bak-Sneppen model. We call this new construction the *forgetful* Bak-Sneppen model, as the exact fitness values will be no longer fixed (or remembered). This idea borrows heavily from the 'locking thresholds representation' in [37], and was used in a much simpler form in Chapter 3 [42]. The forgetful Bak-Sneppen model is defined below and then argued to be equivalent to the normal Bak-Sneppen model, in the sense that at all times, the fitnesses have the same distributions.

Consider a Bak-Sneppen model on a finite transitive graph G with N vertices. To start with, all N vertices have independent uniformly $(0,1)$ distributed fitnesses. In the forgetful Bak-Sneppen model, all N vertices have *fitness distributions*, instead of fitness values. At time 0, all vertices have uniform (0,1) fitness distributions. The system at time n is generated from the system at time $n-1$ by the following procedure.



1. We draw N new independent random variables according to the appropriate fitness distributions at time $n-1$.
2. The minimum of the fitnesses is found and fixed.

3. All the other fitness values are discarded, and replaced by the conditional *distribution* of these fitnesses, given that they are larger than the observed minimal fitness.

4. The vertex with minimal fitness and its neighbours have their value or fitness distributions replaced by uniform (0,1) distributions. (So now all vertices have some distribution associated with them.) This is the state of the system at time n.

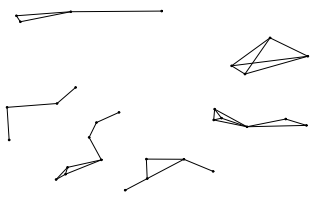
It is easy to see that the fitness distributions at time n generated by this procedure are the same as the fitness distributions in the normal Bak-Sneppen model at time n.

Furthermore, all fitness distributions have the convenient property that they are uniform distributions. Indeed, suppose that a random variable Y has a uniform $(y, 1)$ distribution, denoted by F_y. If we condition on $Y > z$, then Y has distribution $F_{y \vee z}$, where $y \vee z = \max\{y, z\}$. All our fitnesses initially have uniform (0,1) distributions. Two things can change these distributions. They can be reset to F_0 by being updated, or they can be conditioned to be bigger than some given value; in both cases they remain uniform.

The above construction gives a forgetful Bak-Sneppen model on a finite graph, but this is easily extended to a forgetful Bak-Sneppen model on a locally finite graph. The only difference is that initially we assign the fitness distribution F_0 to the origin. The remaining vertices have a fitness distribution with all mass in the point 1, denoted by F_1. The avalanche ends when all the fitnesses within the avalanche are above the threshold, which is equivalent to saying that the minimal fitness is above the threshold. It is possible to see when the avalanche has finished by checking the value of the minimum fitness (phase 2 above). Thus we can use the forgetful method to generate avalanches.

4.3 The construction of the coupling

This section is divided into three parts. To begin with, some intuition behind the main result (Theorem 4.3) is given. This is followed by a precise description of the coupling, and then we give an example for added clarity.



4.3.1 Intuition

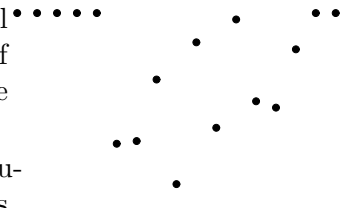
We are interested in comparing the Bak-Sneppen avalanche with the open cluster at the origin of

independent site percolation, with the proviso that the origin is open with probability 1 rather than with probability p. This clearly has no effect on the critical value.

Typically, site percolation is studied as a static random structure, but it is also possible to build up the open cluster at the origin dynamically. This is standard (we refer to [33] for details) but the idea can be described as follows. Starting with just the origin, we can evaluate one of the neighbours and decide whether this neighbour is open or not. If it is, we add it to the cluster, if it isn't, we declare it closed. One can continue in this fashion, each time step evaluating neighbours of the current cluster one by one. If the probability that a vertex is open, given the full history of this process, is always equal to p, then in fact we do create the site-percolation open cluster of the origin. When there are no more unevaluated neighbours, the process stops, and the cluster is finite in that case.

The growth of both a Bak-Sneppen avalanche and the open cluster at the origin is driven by the *extremal* vertices. In a Bak-Sneppen avalanche, the extremal vertices are those vertices that are contained within the avalanche and have neighbours outside the avalanche. It is only through one of the extremal vertices having the minimal fitness that the range of the avalanche can increase. For site percolation, the extremal vertices are those having a neighbour in the open cluster at the origin, but that are themselves unknown as to be open or closed. These are exactly the vertices at the edge of the cluster, and they will increase the size of the cluster by being open. Since it is the extremal vertices that drive the spread of both processes, the task is to relate the two sets of extremal vertices to each other.

The major difficulty to overcome is that in the Bak-Sneppen model an extremal vertex may be updated by neighbouring activity before having minimal fitness itself, whereas in site percolation a vertex is either open or closed. So in the Bak-Sneppen model it is possible that a previously active extremal vertex never has minimal fitness, having been made inactive by a subsequent neighbouring update. Conversely, an originally inactive vertex can be made active. Hence, in the Bak-Sneppen model the neighbour of an active vertex will not necessarily be updated, while in our construction of the open cluster at the origin in site percolation, the neighbour of an open site is always considered. This means that it is not useful to couple the two models in the natural manner by realising the fitness and determining if the vertex is open and closed immediately with the same random variable.



The following heuristics make Theorem 4.3 plausible. If a vertex's fitness is not minimal, then its

conditional distribution based on this information is stochastically larger than its original uniform $(0, 1)$ distribution. So if a vertex is updated by a neighbour having minimal fitness, this makes its fitness stochastically smaller, making the vertex more likely to be active and therefore, intuitively at least, the avalanche is more likely to continue. This means that on average the interference from the non-extremal vertices of the Bak-Sneppen model on the extremal vertices should be beneficial to the spread of the avalanche.

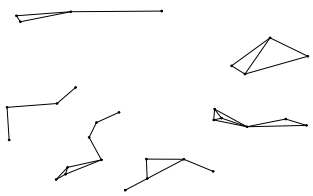
4.3.2 The coupling

We now describe the construction of a process that we will refer to as the *coupled process*. As we shall see later, this process is constructed in such a way that it is stochastically dominated by the Bak-Sneppen avalanche, which is crucial for our argument. In Section 4.5 we show that this coupled process in fact constructs the cluster at the origin of site percolation.

Let $V(G)$ be the vertex set of the graph G. The coupled process is a stochastic process with values in $\{([0, 1] \times \{f, d\})^A, A \subset V(G)\}$. An entry (a, f) means that the value of that vertex is fixed at a forever, while an entry (a, d) means that the value of that vertex is distributed uniformly on $(a, 1)$. The coupled process is coupled to a forgetful Bak-Sneppen avalanche, and is constructed as follows.

Fix an avalanche threshold p. We start with two copies of the graph G, denoted by G_B (for the Bak-Sneppen avalanche) and G_C (for the coupled process). Initially we assign the value 0 to the origin of G_B and $(0, f)$ to the origin of G_C, and we call the origin in G_C *open* (as anticipated before). Then all the Δ neighbours of the origin of both graphs get distribution F_0. On G_C, we define the *extremal set* $\mathcal{E}$ as the set of all points that have been assigned a distribution, but not (yet) an exact value.

The Bak-Sneppen avalanche on G_B is generated according to the aforementioned (forgetful) construction, i.e., we sample new fitnesses, fix the minimal value and then calculate the fitness distributions accordingly. In the coupled process, only the vertices contained in $\mathcal{E}$ are considered. We apply the following procedure to all vertices in $\mathcal{E}$.



Consider a vertex $v_C \in \mathcal{E}$ with G_B-counterpart v_B. Let F_z and F_y be their respective fitness distributions. We realise the fitnesses of the vertices in G_B, and in particular realise the fitness of v_B with an independent uniformly (0,1) distributed random variable U via $y + (1 - y)U$. Let M be the minimal fitness in G_B. As long as the vertex with

minimal fitness in the avalanche is active, i.e., $M \leq p$, we have the following two options, with corresponding rules for the coupled process. One should bear in mind that the main goal of the coupling is the stochastic domination. In Section 4.3.3 below, these are illustrated by an explicit example.

1. The fitness of v_B is not minimal.
   We alter the distribution of v_C by conditioning on the extra information that the fitness of v_B must be bigger than M. Since the fitness of v_B is not minimal, we have $y + (1-y)U > M$, and hence $U > (M-y)^+/(1-y)$. The new distribution of v_C is $F_{\hat{z}}$, where

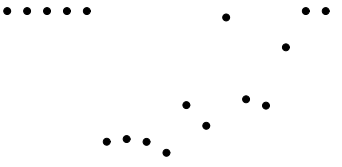
$$\hat{z} = z + (1-z)\frac{(M-y)^+}{(1-y)}.$$

2. The fitness of v_B is minimal, so it has value M.
   It follows that $y + (1-y)U = M$. The fitness of v_C is now *fixed* at

$$z + (1-z)U = z + (1-z)\frac{(M-y)}{(1-y)}.$$

   If this value is less than p, we say that v is *open*, remove v from $\mathcal{E}$, add the neighbours of v that have an undetermined state to $\mathcal{E}$, and give them distribution F_0. If the value of v is larger than p, then v is *closed* and removed from $\mathcal{E}$.

The final step of the construction is as follows. The first time that the vertex with minimal fitness in G_B is inactive (that is, $M > p$), the Bak-Sneppen avalanche has finished. As soon as this happens, we fix *all* the values of the vertices in $\mathcal{E}$ in the following way, similar to rule 2 above. Let $v_C \in \mathcal{E}$ and v_B have fitness distributions F_z and F_y respectively, and let U be the associated uniform (0,1) random variable. Then U satisfies $y + (1-y)U \geq M$, i.e., $U \geq (M-y)/(1-y)$. The new distribution of v_C is $F_{\hat{z}}$ with $\hat{z} = z + (1-z)(M-y)^+/(1-y)$. As final step of the coupling, we realise the fitness of v_C as $\hat{z} + (1-\hat{z})X$, where X is an independent uniformly (0,1) distributed random variable. In Section 4.4 we show that as soon as the Bak-Sneppen avalanche ends, this fitness value is at least p, and hence all the vertices in $\mathcal{E}$ will be closed. Before that, we give an example to illustrate the coupling procedure described above.



The behaviour of the processes is illustrated by the following example, displayed in Figure 4.1. In

this example the graph G is a tree. For illustration purposes, we show only the part of the graph where the activity takes place.

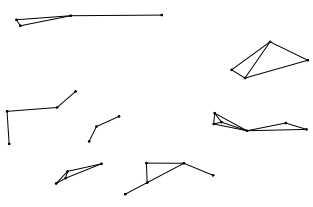
Consider the forgetful avalanche at time n say, in the following situation, see Figure 4.1, graph a: all fitnesses shown have distribution F_0. Then in graph b, uniform $(0,1)$ random variables are drawn. The random variables U_1, U_2, and U_3 are associated with the vertices visible in the picture. Other random variables are of course drawn for the other vertices. We then use the full set of random variables to determine the location and the magnitude of the new minimal fitness. This happens to be the vertex corresponding to the random variable U_2 (graph c). Finally, the new fitness distributions for time $n+1$ are determined (graph d). Note that the forgetful Bak-Sneppen model never actually assumes the values given in graph b.

4.3.3 An example

During the same time step, the coupled process evolves as follows. We only consider the vertices in the extremal set $\mathcal{E}$, see Figure 4.1, graph e. Before the time step, the vertices have fitness distributions F_x, for some $x \in (0,1)$. Given the location of the minimal fitness in the Bak-Sneppen avalanche, the vertices in $\mathcal{E}$ are classified according to the rules 1 and 2 above (graph f). From the location and magnitude of the minimal fitness of the avalanche, it follows that $U_3 \geq M$, so $\hat{x} = x + (1-x)M$. Finally, the value of the vertex that corresponds with the vertex with minimal fitness in the avalanche is fixed, according to rule 2. Its value f is given by $f = x + (1-x)M$. Now there are two possible cases: either $f > p$, and the vertex is closed (graph g), or $f \leq p$, and the vertex is open and its undetermined neighbours are added to $\mathcal{E}$ with distribution F_0 (graph h).

4.4 A domination principle

To show that the critical value of the coupled process can be no smaller than the critical value of the Bak-Sneppen avalanche, we use a domination argument. The propositions below show that the coupled process can finish no later than the Bak-Sneppen avalanche (so that the avalanche can be said to dominate the coupled process).



Proposition 4.6. *For every $v_C \in G_C$ and corresponding $v_B \in G_B$, at all times, the (conditional) fitness distribution of v_C is stochastically larger than the (conditional) fitness distribution of v_B.*

Proof: It should be noted that this proposition only makes sense for vertices in $\mathcal{E}$. Furthermore,

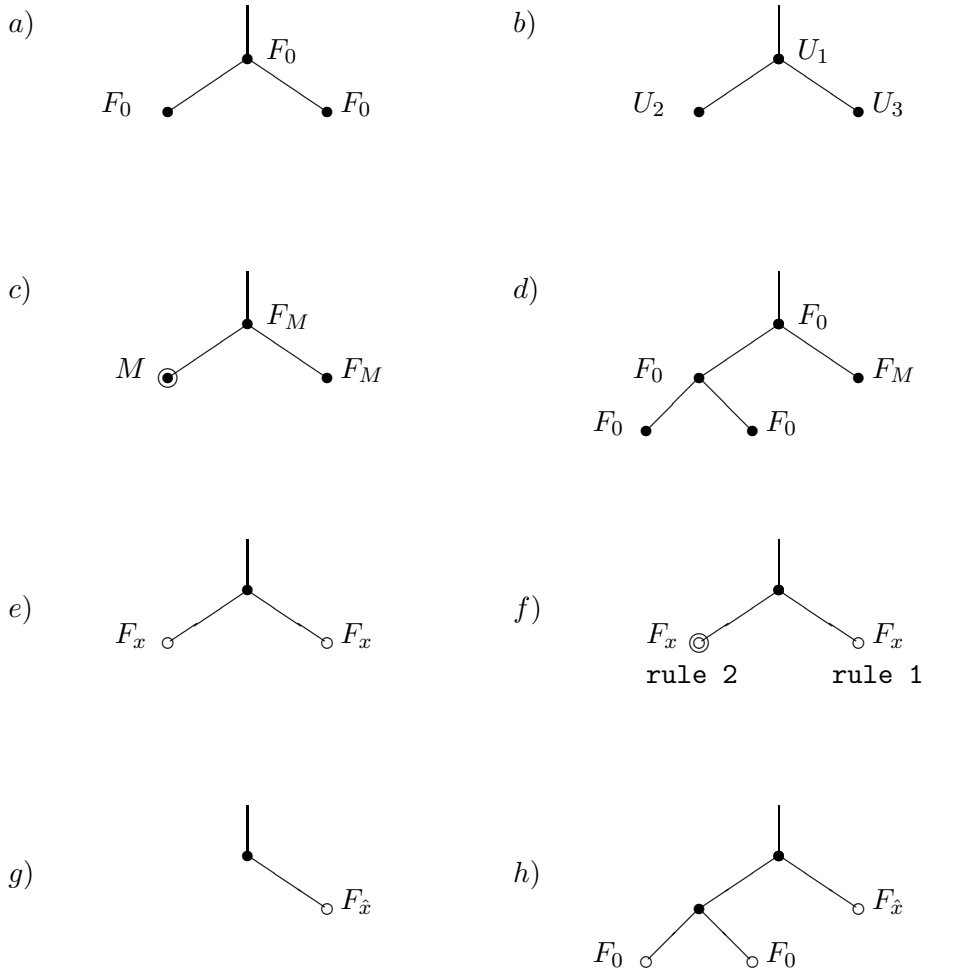
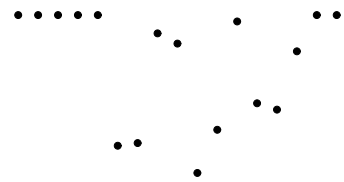


Figure 4.1: A time step in the forgetful Bak-Sneppen process ($a - d$) and in the coupled process ($e - h$). The encircled vertex has the minimal fitness. In the coupled process, black points are open, white points are undetermined, and closed points are omitted.



it is safe to assume that the p-avalanche is still in progress, so the minimal fitness is less than p. The proof proceeds by induction. When new vertices are added to the coupled process, they (and their equivalents in G_B) have uniform $(0,1)$ distributed fitnesses. This is by definition for the coupled process, but also holds for G_B, since vertices in G_B corresponding to new vertices added to $\mathcal{E}$ are always neighbours of the vertex with minimal fitness. This means that the statement of the proposition holds for new vertices added to $\mathcal{E}$.

To make the induction step, consider $v_C \in \mathcal{E}$ with corresponding vertex v_B in G_B, and let F_z and F_y be the fitness distributions of v_C and v_B at time n, where $y \leq z < 1$. Let u be the realisation of the uniform $(0,1)$ random variable associated with v_C and v_B at the intermediate step, and let m be the minimal fitness.

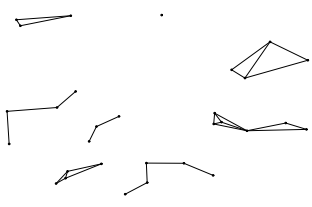
Assume first that v_B does not have the minimal fitness. This provides information on the value of u, namely that $y + (1-y)u > m$, and hence $u > (m-y)/(1-y)$. If $y \geq m$, this information is useless: we already knew that $u > 0$, and the fitness distributions of v_C and v_B are not changed. If $y < m$, then the inequality for u does contain information, and we can calculate the corresponding inequality for the fitness of v_C:

$$z + (1-z)u > z + \frac{(1-z)(m-y)}{1-y} = m + \frac{(1-m)(z-y)}{1-y} := \hat{z}.$$

So at time $n+1$, v_B has distribution $F_{y \vee m}$ and v_C has distribution $F_{\hat{z}}$. Since $m, y < 1$ and $y \leq z$, we have $\hat{z} \geq m$. Hence $(y \vee m) \leq \hat{z}$, and the desired property holds.

Second, we consider the case that a neighbour of v_B had minimal fitness. In that case the fitness distribution of v_B is reset to F_0, and there is nothing left to prove. □

Proposition 4.7. *At the moment that the p-avalanche ends, all vertices in $\mathcal{E}$ are closed. As a consequence, if the probability of an infinite p-avalanche is zero, then there cannot be an infinite cluster of open sites in the coupled process, almost surely.*



Proof: By Proposition 4.6, at all times every point in $\mathcal{E}$ has a fitness that is stochastically larger than the fitness of the corresponding vertex in the avalanche. Hence, if the p-avalanche ends, then in the coupled process all neighbours in the set $\mathcal{E}$ will be closed, as their fixed values can not be smaller than those in the avalanche, which are already greater than p as the avalanche has ended.

This removes all vertices from $\mathcal{E}$ and ensures that no more are added, implying that in the coupled process no more vertices will be added to the open cluster around the origin. □

We conclude this section by giving an example where the coupled process is finite, but the Bak-Sneppen avalanche is infinite. This shows that the stochastic domination described in this section is not a stochastic equality.

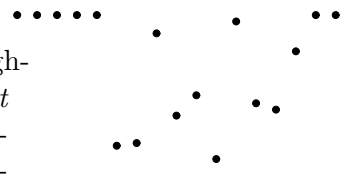
Let $G = \mathbb{Z}$ and $p = 0.7$. Suppose that both in the first step and the second step in the Bak-Sneppen model, the origin is minimal with fitness 0.5. In the coupled process, the neighbours of the origin have fitness distribution $F_{0.5}$ after the first step, and $F_{0.5+(1-0.5)0.5} = F_{0.75}$ after the second step. Since $0.75 > p$, this implies that the neighbours of the origin will eventually be closed, and the cluster in the coupled process is finite. However, the Bak-Sneppen avalanche may very well be infinite.

4.5 The cluster at the origin of site percolation

To complete the proof of Theorem 4.3, it remains to show that the coupled process in fact constructs the open cluster at the origin of independent site percolation, with the proviso that the origin is open with probability 1. To get into the right frame of mind for the proof, we first give an example. At the same time, the example illustrates the construction of the coupled process in action.

4.5.1 An example

Consider the Bak-Sneppen avalanche and the coupled process defined on $\mathbb{Z}$ with parameter p. We wish to calculate the probability that in the coupled process both neighbours of the origin are closed. Note that for the site percolation cluster this probability is $(1-p)^2$, so our aim is to show that this probability is also $(1-p)^2$ for the coupled process. To calculate this probability, we introduce the following, more general probability: for all $0 \leq x \leq p$, let $g_p(x)$ be the probability that both neighbours of the origin will be declared closed, given that their current fitness distributions both are F_x. In this notation, the desired probability is equal to $g_p(0)$.



Starting with the distributions F_x for both neighbours, we call the first subsequent step, the *first* time step. For the coupled process, both neighbours of the origin are declared closed if their realised values are above p. Noting that both neigh-

bours have distribution F_x, this will happen at the first time step if in the Bak-Sneppen model all three values are above $(p-x)/(1-x)$. If the minimum, which has density $3(1-b)^2$, is below $(p-x)/(1-x)$, and located at the origin (which happens with probability $1/3$), then we have to look at subsequent updates in the Bak-Sneppen model.

In this second case, the three fitness distributions in the Bak-Sneppen model are reset to F_0. However, in the coupled process, the fitnesses of -1 and 1 are now $F_{x+(1-x)b}$, where b is the avalanche minimum at the first time step. For the second time step, we are now in a similar situation as for the first, except that the fitness distribution has a different parameter: $x+(1-x)b$ instead of x. This similarity holds for any starting level x, and leads to the following expression for $g_p(x)$:

$$g(x) := g_p(x) = \Big(\frac{1-p}{1-x}\Big)^3 + \frac{1}{3}\int_0^{\frac{p-x}{1-x}} 3(1-b)^2 g(x+(1-x)b)db. \qquad (4.3)$$

Substituting $y = x+(1-x)b$, equation (4.3) becomes

$$g(x) = \frac{1}{(1-x)^3}\Big((1-p)^3 + \int_x^p (1-y)^2 g(y)dy\Big). \qquad (4.4)$$

Using (4.4), a little algebra yields that for small h,

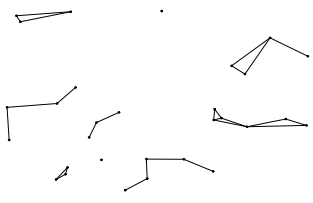
$$g(x+h)-g(x) = \frac{(1-x)^3-(1-x-h)^3}{(1-x-h)^3}g(x) - \frac{1}{(1-x-h)^3}\int_x^{x+h}(1-y)^2 g(y)dy. \qquad (4.5)$$

Since $0 \le g \le 1$, it follows from (4.5) that $g(x+h)-g(x) \to 0$ for $h \to 0$, so g is continuous. Hence, we can calculate the differential quotient:

$$\lim_{d\downarrow 0}\frac{g(x+h)-g(x)}{h} = \frac{3(1-x)^2 g(x)}{(1-x)^3} - \frac{(1-x)^2 g(x)}{(1-x)^3} = \frac{2g(x)}{1-x}.$$

The same holds for the left-hand limit, so $g(x)$ is differentiable, and $g'(x) = 2g(x)/(1-x)$. This differential equation has a unique solution for each p, given by $g(x) = c(p)/(1-x)^2$. Using the boundary condition $g_p(p) = 1$, we find $c(p) = (1-p)^2$, so that

$$g_p(x) = \frac{(1-p)^2}{(1-x)^2}.$$



In particular, the desired probability that in the coupled process both neighbours are closed is given by $g_p(0) = (1-p)^2$, as required.

Although this example gave us what we wanted, clearly this type of calculation does not generalise to more complicated events. Therefore, the proof that the coupled process constructs the site percolation open cluster, which we turn to now, necessarily has a different flavour.

4.5.2 The proof

Our first goal is to determine the distribution of the information we use to generate the coupled process. More precisely, consider an arbitrary step of the forgetful Bak-Sneppen model, when there are n vertices in the avalanche range so far. We enumerate these vertices $1, \ldots, n$, and suppose that all n vertices in the avalanche have just been assigned a (conditional) distribution $F_{y_1}, \ldots, F_{y_n}$. (Recall that these are just uniform distributions above the respective y_i's.) We sample from this random vector, using independent uniform $(0,1)$ distributed random variables $U_1, \ldots, U_n$: a sample from F_{y_i} is realised via $y_i + (1-y_i)U_i$. We locate the minimum M, at vertex K say; note that both M and K are random. Hence,

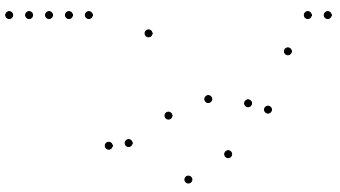
$$U_K = \frac{M - y_K}{1 - y_K}. \tag{4.6}$$

Conditional on K and M, the remaining values U_i, $i \neq K$, are uniformly distributed above $\max\{y_i, M\}$ respectively, that is, we know that

$$U_i > \frac{(M - y_i)}{1 - y_i}, \qquad i \neq K.$$

When we now also sample from all the other entries $i \neq K$, (which are uniformly distributed above $(M - y_i)^+/(1 - y_i)$ respectively) we have described a somewhat complicated way of sampling from the original vector $(U_1, \ldots, U_n)$, that is, such a sample yields independent uniform $(0,1)$ distributed entries, see also Figure 4.2 and its caption. Note that we do not claim that U_K is uniformly distributed on $(0,1)$: it is not. However, since the index K is random, this does not contradict the fact that the vector $(U_1, \ldots, U_n)$ consists of independent uniform $(0,1)$ random variables.

Looking back to Section 3.2, it should be clear that in the coupled process independent uniform $(0,1)$ random variables generated in the above way are used to alter the values of the vertices contained in $\mathcal{E}$. Note that using $|\mathcal{E}|$ entries rather than n does not affect their marginal distributions or dependence structure, as the values of $U_i's$ do not depend on whether the associated vertices are in $\mathcal{E}$ or not.

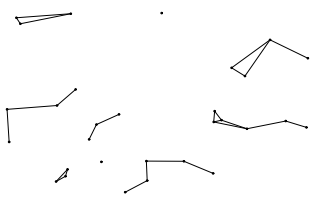


It is now possible to give a direct description of the construction of the coupled process. We start with the origin being open and look at the neighbours of the origin, which initially have distribution F_0. These distributions are realised as follows, using the independent uniform $(0,1)$ random variables described above. At each time step at most one value becomes fixed and the rest are given distributions. The fixed value corresponds to the case that $K \in \mathcal{E}$. To calculate the new values of vertices in $\mathcal{E}\backslash\{K\}$, we use the information that the U_i's are independently and uniformly distributed above $(M-y_i)^+/(1-y_i)$. This means that we do not fix their actual values at that time step, but instead change their distributions conditioned on this information. Once a vertex has a fixed value, it is declared open if and only if this value falls below p. Whenever a vertex is declared open, the neighbours that neither have a fixed value nor belong to $\mathcal{E}$ are added to $\mathcal{E}$ with distribution F_0.

Since fitnesses are initially independent uniform $(0,1)$ when added to $\mathcal{E}$ and the information we use to update the distributions is also independent uniform $(0,1)$, the following holds: if at any time point the procedure is stopped and all the distributions are realised, one will recover an independent uniform $(0,1)$ sample. Hence, all considered vertices (except the origin) are open independently and with probability p. It should now be obvious that our procedure is no different to building a site percolation cluster at the origin by the iterative method of assigning independent uniform $(0,1)$ random variables to all undetermined neighbours of the cluster and calling a vertex open if its random variable takes a value less than p. This completes the proof of Theorem 4.3. □

Note that in case of an infinite Bak-Sneppen avalanche, some vertices in the coupled process may never get a fixed value. This is not a problem, because this is just what happens if an infinite open cluster around the origin is built up dynamically: not all vertices will be tested in the process of constructing this cluster.

4.6 Final remarks and extensions



Throughout this paper we have only considered locally finite transitive graphs. We assumed transitivity to avoid technicalities that would have obscured the main lines of reasoning. However, our results also hold in a more general setting, namely for any locally finite graph. The following observations explain this generalisation. The lower bound

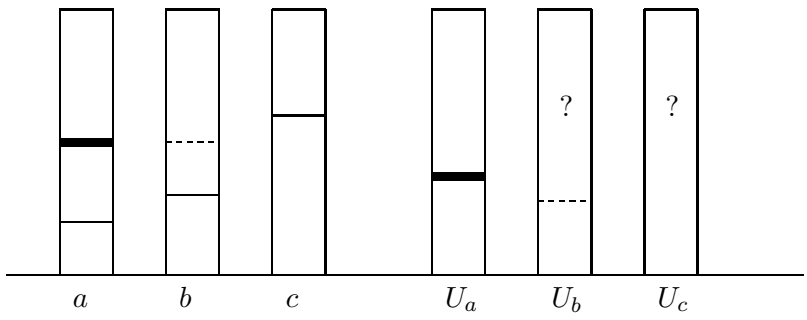
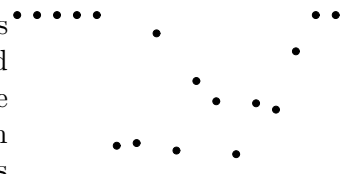


Figure 4.2: In the forgetful BS-avalanche, before the update, vertices a, b and c have fitness distributions $F_{0.2}$, $F_{0.3}$, and $F_{0.6}$, respectively. After realising these distributions, vertex a is minimal with value $M = 0.5$. This means that $U_a = \frac{0.5-0.2}{1-0.2} = \frac{3}{8}$, $U_b \geq \frac{0.5-0.3}{1-0.3} = \frac{2}{7}$, and $U_c \geq 0$. This sample, namely $U_a = 3/8$ combined with a sample from a uniform $(2/7, 1)$ and a uniform $(0, 1)$ distribution, is a sample of three i.i.d. uniform $(0, 1)$ random variables.

(Proposition 4.2) can easily be adapted by considering a branching process with binomial$(\Delta^* + 1, p)$ offspring, where Δ^* is the maximal degree of the graph. Note that the lower bound is trivial if $\Delta^* = \infty$.

The coupling argument used to prove that the Bak-Sneppen avalanche dominates site percolation, at no point used the transitivity of the underlying graph, and hence also holds for non-transitive graphs. However, for non-transitive graphs, the choice of the origin affects the behaviour of the avalanche. The upper bound (Theorem 4.3) is generalised by the following observation: although the distribution of the size of the open cluster around the origin in site percolation does depend on the choice of the origin, standard arguments yield that the critical value does not.

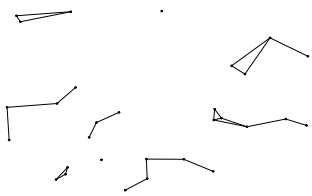
Another consequence of our methods is the following. The careful reader may have noticed that the proofs actually yield a stronger result than stated in Theorem 4.3, namely *stochastic domination*. Define the range of site percolation to be the cardinality of the open cluster around the origin plus all its closed neighbours (these closed neighbours correspond to updated vertices in the Bak-Sneppen avalanche that were never minimal). The proof of Theorem 4.3 then demonstrates that the range of the p-avalanche is *stochastically larger* than the range of site perco-



lation with parameter p.

Although not explicitly stated in the proof of Proposition 4.2, a similar extension also applies there. The set of offspring of a branching process with a binomial $(n-1,p)$ offspring distribution is equivalent to the open cluster around the origin (root) of site percolation with parameter p on T_n^*, where T_n^* is a rooted tree where the root has degree $n-1$, and all other vertices have degree n. In this case we get that the range of a p-avalanche on a transitive graph with common vertex degree Δ is *stochastically smaller* than the range of site percolation on $T_{\Delta+2}^*$.

Finally, we argue that Theorem 4.3 holds as well for the critical value (4.2). It is well-known that for site percolation on $\mathbb{Z}^d$ or on a tree, $p_c^{site}(G)$ is equal to the critical value associated with the expected size of the open cluster at the origin, see Grimmett [33]. Since each vertex in the open cluster contributes at most Δ closed neighbours to the range of site percolation, the range is always less than Δ times the size of the cluster. Hence, the critical values associated with the expectation of these two objects are the same. As a consequence, the stochastic bounds given above imply that the bounds in Proposition 4.2 and Theorem 4.3 also hold for the critical value (4.2).



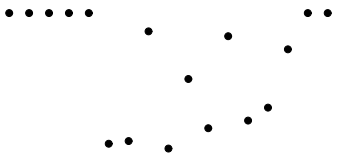
Chapter 5

A near-neighbour continuum percolation model

This chapter introduces a continuum percolation model defined on the points of a d-dimensional homogeneous Poisson process. Each Poisson point is connected to all points within its connection range, which depends on the distances to the other Poisson points. We show that the new model exhibits a phase transition, and obtain results about the critical values in low and high dimensions. This chapter consists of joint work with Nuyens and has been submitted for publication [45].

5.1 Introduction

A continuum percolation model consists of a point process and a rule for connecting the points. The first mention of these types of models appears to date from 1961 [47]. Typically, the point process is a homogeneous Poisson process X on $\mathbb{R}^d$ with some given density λ, and the points are connected in one of two ways. In a *random connection* model, points are connected to each other by undirected edges determined by some probability measure. In a *Boolean* model, a ball is placed around each point in X, and its radius is generated by some probability measure. Two points are considered connected if their balls overlap, and clusters are formed in the obvious way. An overview of these types of models can be found in [51].



In their 1996 paper on continuum percolation [48], Meester and Häggström looked at some models for which the density of the underlying Poisson process is irrelevant to the percolation probability.

In these models, altering the density is equivalent to rescaling the model. In this paper, we generalise this work. Let us begin by reviewing one of their original models.

The *nearest neighbour* model is an example of a random connection model. The model is defined by the following connection rule: connect each $x \in X$ to its k nearest neighbours. Letting U denote the event that an infinite component is formed, we define

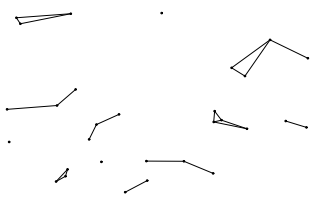
$$k_c(d) = \min\{k \geq 1 : \mathbb{P}(U) > 0\}.$$

By ergodicity, we have that $k_c(d) = \min\{k \geq 1 : \mathbb{P}(U) = 1\}$. It has been proven for this model that $k_c(1) = \infty$ and that $2 \leq k_c(d) < \infty$, for $d \geq 2$, see [48]. Furthermore, it is known that $k_c(d) = 2$ for large d. It is also believed, but unproven, that $k_c(2) = 3$ and $k_c(d) = 2$ for $d \geq 3$. We shall denote the nearest neighbour model by $\mathrm{NN}(d, k)$, where d is the dimension and $k \in \mathbb{N}$ is the parameter.

In the next section we introduce a generalised nearest neighbour model, of which the $\mathrm{NN}(d, k)$ model is a special case. The main questions that we shall study in this new model concern the percolation probability, and in particular the existence of phase transitions as the dimension of the Poisson process (or any other parameter for that matter) is varied. An important tool at our disposal is renormalisation. Renormalisation means discretising the space into boxes to form a lattice and then looking at the connections between neighbouring boxes. This then permits a comparison with site percolation. There are a number of examples of this kind of approach [52, 46]. Some of our proofs will be based on the methods used in [48].

5.2 A generalised nearest neighbour model

Let X be a homogeneous Poisson process on $\mathbb{R}^d$. Consider a point $x \in X$ and let $d_i(x)$ denote the Euclidean distance from x to its ith nearest neighbour. The $\mathrm{NN}(d, k)$ model can now be defined by the connection rule "x is connected to all points within distance $d_k(x)$ of x." We want to generalise this model by not considering the $d_i(x)$ themselves, individually, but a function of these distances. However, we wish to maintain the density invariance of the model. Under the extra assumption that $f(\{d_i(x)\}_{i\geq 1}) = \sum_{i=1}^{\infty} f_i(d_i(x))$, then the $\{f_i\}_{i\geq 1}$ have to be linear maps from $\mathbb{R}^+$ to $\mathbb{R}^+$, i.e., $f_i(x) = \alpha_i x$. This condition gives us a class of functions indexed by the α_i values. If the f_i are allowed to be functions of



more than one of the $d_i(x)$, then non-negative measurable functions of the form $d_i(x)f(\frac{d_j(x)}{d_k(x)})$ also preserve density invariance.

With this motivation, we define a new model by connecting x by an edge to all points y with $|x-y| \leq r(x)$, where $r(x)$ is given by

$$r(x) = \sum_{i=1}^{\infty} \alpha_i d_i(x),$$

for some vector $\underline{\alpha} = (\alpha_1, \alpha_2, \dots)$ with $\alpha_i \geq 0$ for all i. We call this model the *generalised nearest neighbour* model, and denote this model by $\mathrm{GN}(d, \underline{\alpha})$. Since we shall begin our analysis by looking at simpler versions of this model, we denote by $\mathrm{GN}_k(d, \alpha)$ the model when only the kth component of $\underline{\alpha}$ is non-zero, and has value α. Observe that $\mathrm{GN}_k(d, 1) = \mathrm{NN}(d, k)$.

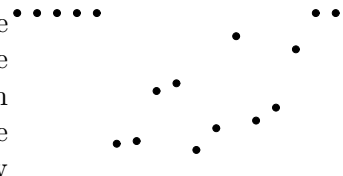
Looking at the $\mathrm{GN}_1(d, \alpha)$ model, we define a critical value for α as before:

$$\alpha_c(d) = \inf\{\alpha \geq 0 : \mathbb{P}(U) > 0\}.$$

Since the critical value for $\mathrm{NN}(d, 1)$ is known to be ∞, we have that $\alpha_c(d) \geq 1$ for all d. One could also consider a corresponding Boolean model, where on each point x of the Poisson process a sphere of radius $\alpha d_1(x)$ is placed.

For the $\mathrm{GN}_k(d, \alpha)$ model, we denote the corresponding critical value by $\alpha_c^{(k)}(d)$. Hence under this notation, $\alpha_c(d) = \alpha_c^{(1)}(d)$. Sometimes, however, we will vary only k and look for a phase transition for $\mathrm{GN}_k(d, \alpha)$ with α and d fixed. Note the following property when $\alpha < 1$: for all points $x \in X$, there are at most $k-1$ points within distance $r(x) = \alpha d_k(x) < d_k(x)$. This leads to the result that the $\mathrm{GN}_k(d, \alpha)$ model is dominated by the $\mathrm{GN}_{k-1}(d, 1)$ model for all $\alpha < 1$. In the $\mathrm{GN}(d, \underline{\alpha})$ model there is no obvious definition of a critical value as we are dealing with a (possibly infinite dimensional) vector. However, one number that will be of interest is $|\underline{\alpha}| = \sum_{i=1}^{\infty} \alpha_i$. Note that percolation is trivial in all dimensions if $|\underline{\alpha}| = \infty$. This case is therefore ignored, which restricts us to $\underline{\alpha}$ such that $\alpha_i \to 0$ as $i \to \infty$.

In the formation of large clusters, there are two phenomena that are competing. First, to form large clusters, points in the cluster should be connected to (many) other points, and to do this, points should be close to each other. On the other hand, points should have a large connection range, so points should be far away from each other. As we shall see in the next section, this makes the model already far from trivial in one dimension. This is in contrast to the $\mathrm{NN}(d, k)$ models, for which it is very easy to show that there is no percolation in one dimension.



Although the generalised nearest neighbour model is a random connection model, it is possible to define a corresponding Boolean model. In the Boolean version, two points are connected if their connection ranges overlap. The proofs of a number of our results can be easily modified to apply to this Boolean model, for example Theorem 5.6 below. Typically any numerical bounds on the critical value, such as Corollary 5.7, will be halved since the connection ranges now combine to make connections.

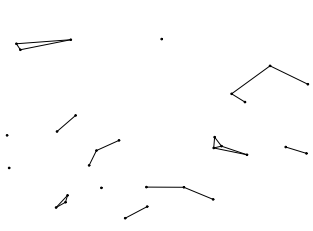
The paper is organised as follows. We first study the GN model in one dimension, in Section 5.3. Then in Section 5.4 we consider the $\mathrm{GN}_1(d,\alpha)$ model, where only the distance to the nearest neighbour is important. In Section 5.5 we study the $\mathrm{GN}_k(d,\alpha)$ model for large d and in Section 5.6 we obtain results for the general $\mathrm{GN}(d,\underline{\alpha})$. The paper concludes with the statement of some open problems.

5.3 The model in dimension 1

For the nearest neighbour model NN(1,k) there is no percolation in one dimension for any k. This result is straightforward and one of the main reasons for this is that if there is no edge between two neighbouring points, then these points belong to separate clusters. For the $\mathrm{GN}_k(1,\alpha)$ model though, this property does not hold and the proof that there is no percolation in one dimension, given below, is non-trivial.

We first introduce some notation. Let $(x, x+m)$ be called a *gap of length* m if $X \cap (x, x+m) = \emptyset$. The term m-gap denotes a gap of length greater than m. We say that there is a *bridge* over a gap if two points on different sides of the gap have an edge between them; we say that x *bridges* a gap if the connection range of x, $r(x)$, spans that gap. Furthermore, a gap $(x, x+m)$ is bridged from the right if there is a point $y \geq x+m$ such that $y - r(y) < x$. The properties of homogeneous Poisson processes imply that for any point $x \in X$, there are infinitely many m-gaps in both the positive (to the right) and negative (to the left) directions for all $m < \infty$.

Let $p(m)$ denote the probability, given that there is an m-gap covering the origin, that this m-gap is not bridged from the right.



Lemma 5.1. *Consider the* $GN_k(1,\alpha)$ *model, and let* $\beta > (\alpha \vee 1)$. *Then* $p(\beta^2) > 0$.

Proof: It will be convenient to use a second distance function, $\bar{d}_k(x)$, where $\bar{d}_k(x)$ is the distance from x to its kth nearest neighbour to the right. Furthermore, call $x \in X$ a β-*point* if $\bar{d}_k(x) > \beta$. Observe that if x is not a β-point, then $r(x) \leq$

$\alpha\bar{d}_k(x) < \alpha\beta < \beta^2$. Hence, when looking for bridges from the right over β^2-gaps, it is sufficient to only consider β points.

Consider a β^2-gap, and the first point to the right of it. Without loss of generality, call this point the origin, 0. Let $0 = X_0 < X_1 < X_2 < \dots$ denote all points to the right of 0 and let $0 \le Y_0 < Y_1 < Y_2 < \dots$ denote the β-points to the right of 0. A sufficient condition for the β^2-gap to be unbridged from the right is that $Y_i - \alpha\bar{d}_k(Y_i) > -\beta^2$ for all $i \ge 0$. From the definition of β-points it follows that $Y_{i+k} - Y_i > \beta$, and hence $Y_i \ge \lfloor \frac{i}{k} \rfloor \beta$.

Since the events $\{\alpha\bar{d}_k(Y_i) \le t\}$ are positively correlated for all i and t, we have

$$\begin{aligned} p(\beta^2) &\ge P(\alpha\bar{d}_k(Y_i) \le Y_i + \beta^2 \text{ for all } i) \\ &\ge P(\alpha\bar{d}_k(Y_i) \le \lfloor \frac{i}{k} \rfloor \beta + \beta^2 \text{ for all } i) \\ &\ge \prod_{i=1}^{\infty} P(\alpha\bar{d}_k(Y_i) \le \lfloor \frac{i}{k} \rfloor \beta + \beta^2). \end{aligned} \tag{5.1}$$

Let Γ have a gamma$(k,1)$ distribution, and note that $\bar{d}_k(x) \stackrel{d}{=} \Gamma$. Since Y_i is a β-point, and $\beta > (\alpha \vee 1)$, we have the following for all $i > k\beta$:

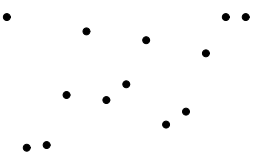
$$\begin{aligned} P(\alpha\bar{d}_k(Y_i) \le \lfloor \frac{i}{k} \rfloor \beta + \beta^2) &\ge P(\bar{d}_k(Y_i) \le \lfloor \frac{i}{k} \rfloor + \beta) \ge P(\bar{d}_k(Y_i) \le \frac{i}{k}) \\ &\ge P(\Gamma \le \frac{i}{k} \mid \Gamma > \beta) = 1 - P(\Gamma \ge \frac{i}{k} \mid \Gamma > \beta). \end{aligned} \tag{5.2}$$

We now use that for $n \to \infty$,

$$P(\Gamma \ge n | \Gamma > \beta) = \frac{\sum_{i=1}^{k-1} e^{-n} \frac{n^i}{i!}}{\sum_{i=1}^{k-1} e^{-\beta} \frac{\beta^i}{i!}} \sim c(\beta, k) n^{k-1} e^{-n}, \tag{5.3}$$

where $1/c(\beta,k) = \sum_{i=1}^{k-1} e^{-\beta}(k-1)!\beta^i/i!$. Combining (5.1), (5.2) and (5.3), we find that $p(\beta^2) > \prod_{i=1}^{\infty}(1-a_i)$ with $a_i \sim c'(\alpha,\beta,k) i^{k-1} e^{-i}$ as $i \to \infty$ and $c'(\alpha,\beta,k) > 0$, and $a_i > 0$ for all i. Since $a_i \to 0$ exponentially fast, we conclude that $p(\beta^2) > 0$. □

Theorem 5.2. *In the $GN_k(1,\alpha)$ model, we have $\alpha_c^{(k)}(1) = \infty$.*



Proof: We prove the theorem by showing that there exist a.s. infinitely many unbridged β^2-gaps in both directions. By symmetry, the probability

that a β^2-gap is bridged from the left is $p(\beta^2)$ as well. Now note that for a gap of given length, between two Poisson points x and y say, the event that it is unbridged from the right is positively correlated with the event that it is unbridged from the left. Indeed, the absence of a bridge from the right makes the distances of points to the right of y to their kth nearest neighbour stochastically smaller, and as a consequence, the same holds for the distances from the points to the left of the gap to their kth nearest neighbour. Therefore, the probability that a β^2-gap is unbridged is at least $p(\beta)^2$, which is strictly positive by Lemma 5.1. Thus there exists an unbridged β^2-gap with positive probability.

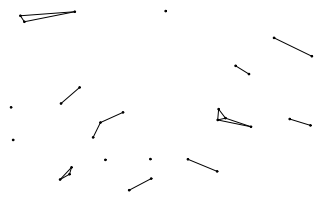
By the ergodicity of the Poisson process, we know that translation invariant events have probability 0 or 1. Since the probability of an unbridged β^2-gap is positive, such a gap exists almost surely. Assume that there is a right-most such gap. By translation invariance such a gap must be uniformly distributed on $\mathbb{Z}$. Such a distribution does not exist, therefore by contradiction there are infinitely many unbridged gaps to the right. This holds similarly to the left and so there is no infinite cluster. □

We now show that if the α_i are large enough, then the GN$(1, \underline{\alpha})$ model does percolate. In fact, we show that for these α_i, the model is *fully connected.*

Proposition 5.3. *In the GN$_k(1, \alpha)$ model, let x be a point of the Poisson process on $\mathbb{R}$. If $\sum_{i=1}^{\infty} i\alpha_i = \infty$, then $r(x) = \infty$ a.s. As a consequence, the GN$(1, \underline{\alpha})$ model is fully connected.*

Proof: First note that $(d_1(x), d_2(x) - d_1(x), \ldots) \stackrel{d}{=} (U_1, U_2, \ldots)$, where $U_1, U_2, \ldots$ are i.i.d. exponential random variables with parameter 2. Hence, defining $d_0(x) = 0$, we may write

$$r(x) = \sum_{i=1}^{\infty} \alpha_i d_i(x) = \sum_{i=1}^{\infty} \alpha_i \sum_{j=1}^{i} d_j(x) - d_{j-1}(x) \stackrel{d}{=} \sum_{i=1}^{\infty} \alpha_i \sum_{j=1}^{i} U_j$$
$$= U_1 \sum_{i=1}^{\infty} \alpha_i + U_2 \sum_{i=2}^{\infty} \alpha_i + \cdots = \beta_1 U_1 + \beta_2 U_2 + \cdots ,$$



where $\beta_k = \sum_{i=k}^{\infty} \alpha_i$. Before continuing, note that $\sum_i \beta_i = \sum_i i\alpha_i$. Denote the Laplace transform of a random variable X by ϕ_X. Since the U_i are independent, the Laplace transform of $V = \sum_i \beta_i U_i$ satisfies

$$\phi_V(s) = \prod_{i=1}^{\infty} \phi_{U_i}(\beta_i s) = \prod_{i=1}^{\infty} \frac{1}{1+\beta_i s/2} = \prod_{i=1}^{\infty} \Big(1 - \frac{\beta_i s}{2+\beta_i s}\Big).$$

Since the U_k are independent and $\{V = \infty\}$ is a tail event, its probability is either 0 or 1, by a Zero-One Law. Hence, $V = \infty$ if and only if $\phi_V(s) = 0$ for all $s > 0$. But $\phi_V(s) = 0$ if and only if $\sum_i \beta_i = \infty$. Hence, if $\sum_i \beta_i = \infty$, then $V = \infty$ a.s., and therefore $r(x) = \infty$ a.s. □

There is an obvious potential extension of this result to higher dimensions. This result is given in Section 5.6.

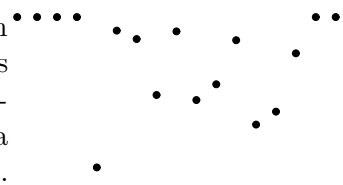
Observe that $\mathbb{E}(r(x)) = \mathbb{E}(U_1)\sum_{i=1}^{\infty}\alpha_i + \mathbb{E}(U_2)\sum_{i=2}^{\infty}\alpha_i + \cdots = \frac{1}{2}\sum_{i=1}^{\infty} i\alpha_i$. Therefore, if the conditions of Proposition 5.3 are not satisfied, the expected connection range is finite. Furthermore, it can be shown that the second moment of $r(x)$ can be bounded from above by a polynomial of its first moment. Therefore, in this case, the variance is also finite. For related one-dimensional independent percolation models, we have the result that there is no percolation if $\mathbb{E}(r(x)) < \infty$ [50]. Thus, it is no surprise that the converse of Proposition 5.3 is true. We begin with a preliminary lemma.

Lemma 5.4. *Let $\sum_{i=1}^{\infty} i\alpha_i < \infty$, then for almost every homogeneous Poisson point process, X, and any $\delta > 0$, there exists an $\epsilon > 0$ and $N < \infty$ (depending on X) such that if N points were added to an interval of length ϵ containing the origin, then all of these points would have a connection range less than δ.*

Proof: Fix X and let $d_i^\star(x)$ be the distance to the i-th nearest neigbour of x in $[1, \infty)$. Let $r^\star(x) = \sum_{i=1}^{\infty} d_i^\star(x)\alpha_i$ and extend the definitions of $r(x)$ and $r^\star(x)$ to $\mathbb{R}$ in the obvious way. Clearly $r(-1) < r^\star(-1) < \infty$. Since $r^\star(-1) < \infty$, there exists an N such that $\sum_{i=N+1}^{\infty} \alpha_i d_i^\star(-1) < \frac{\delta}{2}$. Consider adding N points into a subinterval of length ϵ of $(-1, 1)$. For any added point x, $\sum_{i=N+1}^{\infty} \alpha_i d_i(x) < \frac{\delta}{2}$. Let $\alpha^\star = \max_{i \le N} \alpha_i$, then $\sum_{i=1}^{N} \alpha_i d_i(x) \le N\epsilon\alpha^\star$ for any such point. Therefore $r(x) < \delta$ for $\epsilon \le \frac{\delta}{2N\alpha^\star}$. □

Theorem 5.5. *If $\sum_{i=1}^{\infty} i\alpha_i < \infty$, then all clusters in the GN$(1, \underline{\alpha})$ model are finite almost surely.*

Proof: Recall from the proof of Theorem 5.2 that it is sufficient to show that the origin is in an unbridged gap with positive probability. This follows from a coupling which is a continuum version of the so-called local modifyer [49]. Let Z be a positive random variable with unbounded support. Conditional on Z, let X and Y be homogeneous



Poisson processes with rate λ such that they are independent on $[-Z, Z]$, but identical outside this region. Since the expected connection range is finite, by ergodicity there are only finitely many points whose connection range overlaps the origin. As X and Z are independent and Z has unbounded support, there is a positive probability that $[-Z, Z]$ contains all these points. Assume this happens. Let $D_{N,\epsilon}$ be the event that

$$\begin{cases} X \cap [-Z, -Z+\epsilon] = \emptyset \\ X \cap [Z-\epsilon, Z] = \emptyset \\ |X \cap [-Z, Z]| \leq N \\ |Y \cap [-Z, -Z+\epsilon]| = N \\ |Y \cap [Z-\epsilon, Z]| = N \\ Y \cap [-Z+\epsilon, Z-\epsilon] = \emptyset. \end{cases}$$

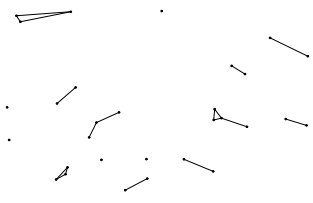
This event has positive probability for any $N > 0$ and $\epsilon \in (O, Z)$. Clearly no point in $Y \backslash [-Z, Z]$ can get a connection range larger than the corresponding point in X. In this case none of them will overlap the origin. By Lemma 5.4, if ϵ is small enough and N is large enough, none of the points of Y in $[-Z, Z]$ will overlap the origin. Therefore the origin will not be bridged by any point in Y. □

Readers familiar with long-range percolation may also see superficial similarities with the one-dimensional homogeneous case where $q(n) < 1$ denotes the probability of being connected to a vertex at distance n. If $\sum nq(n) < \infty$, there is no percolation [33].

5.4 The $\mathbf{GN}_1(d, \alpha)$ model for $d \geq 2$

Having completed our treatment of the special case $d = 1$, we continue by looking at $d \geq 2$, starting with the simplest model: $\mathrm{GN}_1(d, \alpha)$. Recalling that $\alpha_c(d) \geq 1 \ \forall d$, the next result shows that there exists a non-trivial critical value for all $d > 1$.

Theorem 5.6. *For $d \geq 2$, $\alpha_c(d) < \infty$.*



Proof We first consider $d = 2$, with the density of the Poisson process equal to 1. A 3×3 box is called a *banana box* if the 1×1 box in its centre contains exactly one point, and the rest of box is empty, see Figure 5.1. The probability that a 3×3 box is a banana box is $e^{-1}e^{-8}$. A $3n \times 3n$ box is

called *good* if it contains at least one banana box. The probability that a $3n \times 3n$ box is good is at least $1 - (1 - e^{-9})^{n^2}$.

Now choose n so large that the probability that a $3n \times 3n$ box is good is larger than p_c^{site}. Consider two neighbouring good boxes. By construction, in a good box there is a point whose nearest neighbour is at distance at least 1. On the other hand, two points in two neighbouring $3n \times 3n$ boxes are at most $3n\sqrt{5}$ away from each other. Then for all $\alpha > n\sqrt{45}$, any two neighbouring good boxes will be connected to each other. Since the probability that a box is good is larger than p_c^{site}, there is a.s. an infinite cluster of good boxes. Hence, $\alpha_c(2) \leq n\sqrt{45} < \infty$. For $d \geq 3$, the proof is similar, and is therefore omitted. □

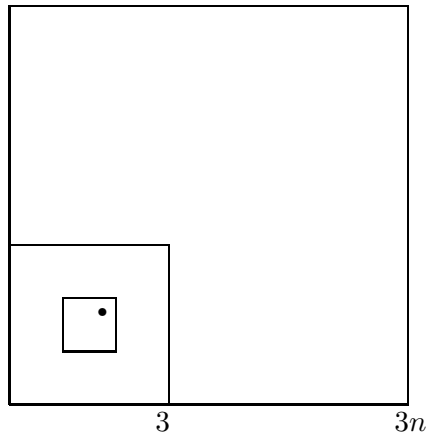


Figure 5.1: A banana box inside a good box of side length $3n$.

The same idea can be used to find a not so sharp upper bound for $\alpha_c(2)$, by optimising the box size used in the proof of Theorem 5.6. Combining this with the best available rigorous upper bound: $p_c \leq 0.679492$ [54] for site percolation on the square lattice, we get the following corollary.

Corollary 5.7. *We have* $\alpha_c(2) < 41$.

We now move to the asymptotical behaviour of $\alpha_c(d)$ for $d \to \infty$.

Theorem 5.8. *We have* $\alpha_c(d) \to 1$ *as* $d \to \infty$.

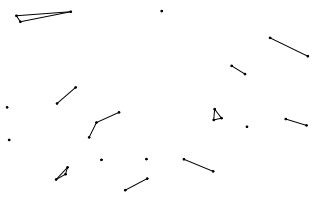
The proof of Theorem 5.8 builds on the work of Meester and Häggström [48] and follows a similar approach to Section 3 of their paper. They varied

k rather than α and looked at the critical value of k for the $\mathrm{GN}_k(d,1)$ model. In their paper, it was proven that $k=2$ is the critical value when d is large. It should be noted that the result was not originally presented in this form and has been expressed here in the terminology of this paper. The technique they use is showing that there is percolation in the $\mathrm{GN}_2(d,1)$ model for all sufficiently large d. Combining this with an earlier result that the $\mathrm{GN}_1(d,1)$ model doesn't percolate for any d yields the result. Here, we derive Theorem 5.8 by showing that for any $\alpha > 1$ the $\mathrm{GN}_1(d,\alpha)$ model percolates for d large enough.

The proof of this result requires a number of steps. Before giving the formal proof, we give an outline of the proof and some preliminary results. Our approach will be to construct a point process that with non-zero probability gives a subset of the cluster at the origin of $\mathrm{GN}_1(d,\alpha)$ that is infinite. The existence of such a process would implies that the $\mathrm{GN}_1(d,\alpha)$ model percolates. The point process is built up by a sequence of steps. A step is said to be *successful* if it produces a (finite) subset of the cluster at the origin of $\mathrm{GN}_1(d,\alpha)$ and initiates two subsequent steps. The idea is that by taking the dimension very large, we can uniformly bound from below the probability that a step is successful. By making this lower bound sufficiently large, we then show that an infinite sequence of steps occurs with positive probability.

The *out-cluster* of a point x in $\mathrm{GN}_1(d,\alpha)$ is the set of points defined by the following iterative procedure, initiated by the set $\{x\}$. Given that we have a set $\{x_1, x_2, \ldots, x_n\}$, add to this set all Poisson points y such that $d(x_i, y) < \alpha d_1(x_i)$ for some i. This is then repeated with the new set of points. The procedure either continues forever, in which case the out-cluster is infinite, or stops when there are no more points to add. By construction, all points in the out-cluster at x must belong to the cluster at x. We shall prove percolation for the $\mathrm{GN}_1(d,\alpha)$ model by showing that (a subset of) the out-cluster at the origin is infinite with positive probability.

We attempt to construct a subset of the out-cluster at the origin via a point process. This point process is built up algorithmically using d-dimensional spatial branching processes. Initially, the space ($\mathbb{R}^d$) is empty and the spatial branching processes place points in this space to create the point process. Each spatial branching process (SBP) is run for n generations and is thought of as a step in the algorithm. A step of the algorithm is declared *successful* if two things happen. First, the union of the SBP with all previously successful steps must form a subset of the out-cluster at the origin of $\mathrm{GN}_1(d,\alpha)$. This means that for all



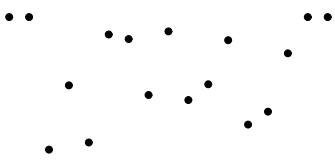
points x in the point process, the offspring of that point must be within distance $\alpha d_1(x)$ of x. Also, the collection of points must obey the law of a homogeneous Poisson process. Second, there must be two points in the n-th generation of the SBP that satisfy a certain location condition as defined below. These two points are used to initiate two more steps of the algorithm. The point process is then defined as the union of all the successful steps of the algorithm.

The SBP lives in $\mathbb{R}^d$, but one can consider its projection onto $\mathbb{R}^2$ via the linear map

$$L(x_1, \ldots, x_d) = \sqrt{d}(x_1, x_2). \tag{5.4}$$

This permits a comparison with oriented site percolation on the lattice $\mathcal{L} = \{(i, j) \in \mathbb{Z}^2 : i \geq 0, |j| \leq i, (i + j)/2 \in \mathbb{Z}\}$, with oriented edges from (i, j) to $(i + 1, j \pm 1)$. Each site (i, j) of $\mathcal{L}$ corresponds to a square $S_{i,j} = [M(i - \frac{1}{2}), M(i + \frac{1}{2})] \times [M(j - \frac{1}{2}), M(j + \frac{1}{2})]$, where M will be chosen later. The second condition for a step to be successful (mentioned above) is now that the projected SBP starting in box $S_{i,j}$ has points in its n-th generation in $S_{i+1,j-1}$ and $S_{i+1,j+1}$. This allows us to start a pair of new SBPs from these points with projected origins in $S_{i+1,j-1}$ and $S_{i+1,j+1}$ respectively. In this way, a successful step in the algorithm corresponds to an open site in $\mathcal{L}$. Let $p_c < 1$ be the critical value of oriented site percolation on $\mathcal{L}$. If for all sufficiently high dimensions we can bound uniformly (i.e., irrespective of what has happened in previous steps of the algorithm) from below the probability of a successful step of the algorithm by some $p > p_c$, , then we have shown that the $\mathrm{GN}_1(d, \alpha)$ model percolates. Before proving this result, we define the spatial branching process and give some properties.

The spatial branching process. (SBP) Let $S_r(x)$ denote the hyper-sphere of radius r centred at x, write $S_r = S_r(0)$, and let $|S_r|$ denote the volume of S_r. The spatial branching process with origin 0 is defined as follows. We start with $Z_0 = \{0\}$. Given Z_n, the offspring of each $y \in Z_n$ is generated by the following procedure. Let $\delta_1 > 0$. An independent homogeneous Poisson point process $X^{(y)}$ on $S_{1+\delta_1}(y)$ is generated, with density $\lambda(d)$ such that the expected number of points contained in $S_1(x)$ is 1. Then a ball is grown around y until either the ball has radius $1 + \delta_1$ or c_2 points of $X^{(y)}$ have been encountered, for a certain $c_2 \in \mathbb{N}$. The offspring of y are then the points of $X^{(y)}$ that are contained in this ball.



In general, the SBP does not generate a homogeneous Poisson process. Consider two points x and y such that $d(x, y) < 1 + \delta_1$. It is clear that

the overlap $S_{1+\delta_1}(x) \cap S_{1+\delta_1}(y)$ is non-empty, and that considering the offspring of both x and y together, the density of points is doubled on this overlap. However, if we condition on y having no offspring on this overlap, then the joint collection of offspring does form a homogeneous Poisson process $S_{1+\delta_1}(x) \cup S_{1+\delta_1}(y)$. Thus if the overlap is small, then with high probability, namely when y has no offspring on this overlap, we can consider the union of the offspring to be a homogeneous Poisson process on $S_{1+\delta_1}(x) \cup S_{1+\delta_1}(y)$. This reasoning will be crucial when comparing the SBP to the out-cluster at the origin of the $\mathrm{GN}_1(d, \alpha)$ model. The following standard result shows that in high dimensions, this overlap is negligible.

Lemma 5.9. *If x_1 and x_2 are such that $d(x_1, x_2) > 0.9$ and $r_1, r_2 \in (0.9, 1.1)$, then*

$$\frac{|S_{r_1}(x_1) \cap S_{r_2}(x_2)|}{|S_{r_1}(x_1)|} \to 0$$

as $d \to \infty$.

Observe that for all $\epsilon > 0$, $|S_{1+\epsilon}|/|S_1| \to \infty$ as $d \to \infty$. The offspring distribution for each individual in the SBP is *distributed like* $Y \wedge c_2$, where Y is Poisson distributed with parameter $|S_{1+\delta_1}|/|S_1|$. Hence, for any $0 < c_1 < c_2$, we can choose $\delta_1 = \delta_1(d)$ such that for d sufficiently large, $\mathbb{E}(Y \wedge c_2) = c_1$ and that $\delta_1(d) \to 0$ as $d \to \infty$. Note that c_1 and c_2 do not vary with the dimension. Thus with this choice of $\delta_1(d)$, the offspring of a point x converge weakly (as $d \to \infty$) to a set of points uniformly distributed on the surface of the unit sphere centred at x. If $c_1 > 1$, then the SBP is supercritical and the probability of extinction can be made arbitrarily small by taking c_1 sufficiently large (independent of the dimension).

The following lemma comes from [53], and used the map L defined in (5.4).

Lemma 5.10. *Suppose $U = (U_1, \ldots, U_d)$ is uniform on the surface of S_1. Then, as $d \to \infty$, the two-dimensional random vector $L(U)$ converges in distribution to the bivariate normal distribution $N(0, I)$ with zero mean and as covariance matrix the identity matrix I.*

The mapping L can be used to map our SBP onto $\mathbb{R}^2$. Let SBP* denote the limit as $d \to \infty$ of the mapped SBP. Since L is continuous, this limit is the same as taking the mapping the limit of the SBP. Thus, by the previous discussion and Lemma 5.10, SBP* is the process where each point has a Poisson(c_1) distributed number of offspring

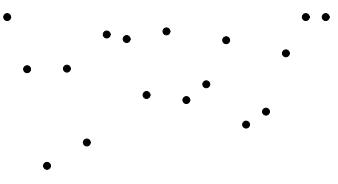
and these offspring are distributed bivariate normally with zero mean and the identity covariance matrix. The same process started at x rather than the origin is denoted by SBP_x^*.

Lemma 5.11. *Given $\epsilon > 0$ and c_1 sufficiently large, we can find a positive integer N_0 and a positive number M, such that for all $x \in S_{i,j} = [M(i-\frac{1}{2}), M(i+\frac{1}{2})] \times [M(j-\frac{1}{2}), M(j+\frac{1}{2})]$, the probability that the N_0th generation of SBP_x^* contains at least one point in $S_{i+1,j-1}$ and at least one point in $S_{i+1,j+1}$ exceeds $1-\epsilon$.*

This result is a small perturbation of a result from [48] and is stated without proof. The process considered in [48] is a branching random walk, so it is (slightly different) than the process considered here. However, the two processes are very similar and the key difference between the branching random walk and our SBP* is that the branching random walk can never go extinct. The role of c_1 in Lemma 5.11 is to make the probability of survival for N_0 generations sufficiently large, so that the result of the lemma applies to SBP* also. We are now ready to prove the main theorem of this section.

Proof of Theorem 5.8: Fix $\alpha > 1$ and consider defining the SBP for a fixed dimension by choosing $\delta_1 > 0$ as follows. First set $\delta_2 > 0$ such that $1 - \delta_2 > \alpha^{-1}$, and then choose δ_1 such that $1 + \delta_1 < (1-\delta_2)\alpha$. Consider running the SBP for n generations and that all points are at least distance $1+\delta_1$ from all others points that aren't its parent. Then, for all $x \in$ SBP, the only points in $S_{1+\delta_1}(x)$ are the offspring of x coming from the homogeneous Poisson process on $S_{1+\delta_1}(x)$ used to generate them. Thus, in this case, the SBP creates a homogeneous Poisson process on the space $\bigcup S_{1+\delta_1}(x)$, where the union is over all points x in the first $n-1$ generations of the SBP. Furthermore, consider that for all such x, the offspring are born at least distance $1-\delta_2$ away, i.e., all offspring of x appear on the annulus $S_{1+\delta_1}(x) - S_{1-\delta_2}(x)$. This implies that $d_1(x) > 1-\delta_2$ for all points x in SBP. Therefore, $r(x) > (1-\delta_2)\alpha > 1+\delta_1$, for all x in SBP.

We generate an object that is dominated by the cluster of the site percolation process by running the following algorithm. The algorithm consists of the steps $(0,0), (1,-1), (1,1), (2,-2), \ldots$, with step (i,j) only being carried out if at least one of the steps $(i-1,j-1)$ and $(i-1,j+1)$ is successful. Step (i,j) consists of a SBP started from a point in $S_{i,j}$ and is called *successful* unless one of the following errors occur.



(a) The spatial branching process fails to reach the two neighbouring boxes in the projected space.

(b) An individual is born within distance $1 - \delta_2$ of its parent.

(c) An individual in the projected SBP is further than R_0 from the origin (of the branching process). The choice of R_0 is given below.

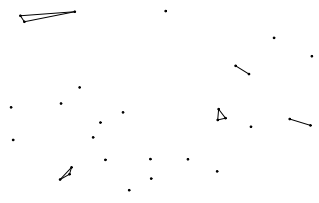
(d) An individual is born within distance $1+\delta_1$ of an already generated individual that is not its parent.

A step is stopped if any of the above errors occur. To avoid ambiguity, step (i, j) is started from the point in generation N_0 of step $(i - 1, j - 1)$ (if successful) or step $(i - 1, j + 1)$ (if successful) whose projection is closest to (Mi, Mj). We let $\mathcal{F}_{i,j}$ denote the σ-algebra generated by the indicator functions of the successes of steps $(0, 0), (1, -1), (1, 1), \ldots, (i, j - 2)$ of the algorithm.

This algorithm generates a point process consisting of all the points contained in all the steps. This point process is then thinned by removing any points that are born in a previously explored part of space. Note that these points are a subset of those that cause type (d) errors and can only occur in unsuccessful steps of the algorithm. This thinning procedure ensures that the point process is a homogeneous Poisson process on a random subset of $\mathbb{R}^d$. Furthermore, we claim that if the probability of a successful step is sufficiently large, this algorithm will with positive probability generate a subset of the $\mathrm{GN}_1(d, \alpha)$ model that contains an infinite cluster.

It should be clear from the above discussion that, taken on its own, a successful step generates a homogeneous Poisson process on a random subset of $\mathbb{R}^d$. Furthermore, by the law of $\mathrm{GN}_1(d, \alpha)$, the origin of the step is contained in the same cluster as two points that can be used to start two subsequent steps. This still holds after the point process is thinned (since the thinning can only affect points in unsuccessful steps). Furthermore, subsequent steps can only interfere with an earlier successful step with precisely those points that are removed with the thinning. Thus, Theorem 5.8 follows from the following claim. For all sufficiently large d, and for all $(i, j) \in \mathcal{L}$,

$$\mathbb{P}(\text{step } (i,j) \text{ is successful} | \mathcal{F}_{i,j}) = p > p_c.$$



The approach is to show that we can simultaneously make the probabilities of each type of error arbitrarily small by suitable parameter choices. The reason for introducing error (c) is to help

bounding error (d) independently of the history of the algorithm. Let γ be such that $1 - 4\gamma > p_c$. We proceed by bounding the probability of each type of error from above by γ.

Lemma 5.11 shows that we can make the probability of error (a) arbitrarily small, for some c_1, N_0 and for all d sufficiently large. We choose c_1 and N_0 such that $\mathbb{P}(\text{type (a) error}) < \gamma$ for all suitably high dimensions. We also fix $1 - \delta_2 > \alpha^{-1}$ and $c_2 > c_1$. Recall from the prior discussion about the spatial branching process that for c_1 and c_2 fixed there exists a $\delta_1(d)$ that defines the required branching process. Therefore, we have also fixed $\delta_1(d)$. Furthermore, since $\delta_1(d) \to 0$ as $d \to \infty$, $1 + \delta_1(d) < (1 - \delta_2)\alpha$ in high dimensions, as required.

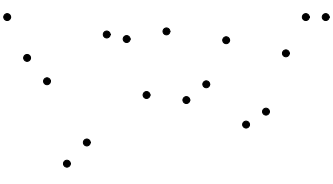
Recall that for all $\epsilon > 0, |S_{1-\epsilon}|/|S_1| \to 0$ as $d \to \infty$. Since the maximum number of points in a step of the algorithm is now bounded by $c_2^{N_0}$ and δ_2 is fixed, we make $\mathbb{P}(\text{type (b) error})$ smaller than γ by taking the dimension high enough. Next we choose R_0 such that the probability that all individuals of SBP* are within distance R_0 of the origin is at least $1 - \gamma$. This implies that in high enough dimensions, the probability of a type (c) error is also less than γ.

It now only remains to bound the probability of a type (d) error. This is the only type of error that depends on the history of the process and thus bounding this error is more involved. We begin by considering a type (d) error in step (i, j) caused by an individual from step $(\hat{i}, \hat{j})$ such that $M\sqrt{(i - \hat{i})^2 + (j - \hat{j})^2} < T_0$, where T_0 is a constant that will be chosen later on. This is straightforward for any T_0 since the number of steps $(\hat{i}, \hat{j})$ to consider is bounded, meaning that the total number of Poisson points is also bounded. Thus, by Lemma 5.9, taking the dimension high enough makes the probability of this type of error less than $\gamma/2$.

To finish the proof, we consider a type (d) error due to a previous step $(\hat{i}, \hat{j})$ satisfying

$$\lfloor M\sqrt{(i - \hat{i})^2 + (j - \hat{j})^2} \rfloor = Q > T_0.$$

We define the *volume* of the step $(\hat{i}, \hat{j})$ to be $|\cup S_{1+\delta_1}(x)|$, where the union is over all points x contained in the step. Note that the volume of step $(\hat{i}, \hat{j})$ is uniformly bounded and that the step (i, j) can only fail because of a type (d) error with step $(\hat{i}, \hat{j})$ if one of its points falls within the volume of $(\hat{i}, \hat{j})$. Note that if an error of type (c) occurs, we stop running the step of the algorithm. So any point that is born a projected distance greater than R_0 from the origin of the step has no offspring and is



not scanned around. Thus, any point in step $(\hat{i}, \hat{j})$ that we scan around must be at least distance $Q - 2R_0$ from any point in step (i, j). Thus from Lemma 5.10 and the exponential decay of the normal distribution, for d large enough, the fraction of the projected volume that falls into the circle with radius R_0 centred at (Mi, Mj) is less than $|S_{1+\delta_1}|/Q^3$ for all large Q.

Since the number of points $(\hat{i}, \hat{j})$ such that $\lfloor M\sqrt{(i-\hat{i})^2 + (j-\hat{j})^2} \rfloor = Q$ is bounded by a constant times Q, and the series $\sum_{q>T_0} q^{-2}$ converges, we can make the total of such volume from all points small by choosing T_0 sufficiently large. In particular, we can make this volume so small that the probability of each individual in step (i, j) coming within distance $1 + \delta_1$ of another point already generated (other than its parent) is less than $\gamma/(2c_2^{N_0})$. Since there are at most $c_2^{N_0}$ points in step (i, j), we get the desired bound.

□

5.5 Results for $\mathbf{GN}_k(d, \alpha)$ for $d \geq 2$

In low dimensions, the difference between GN_1 and GN_k with $k > 1$ can be quite pronounced. However, in high dimensions this is no longer the case, since there the difference between $d_k(x)$ and $d_1(x)$ disappears. This leads to some interesting behaviour.

First, we discuss an immediate consequence of Theorem 5.8. Recall from Section 5.2 that for all $\alpha < 1$, the $\mathrm{GN}_2(d, \alpha)$ is dominated by the $\mathrm{GN}_1(d, 1)$ model and that there is no percolation in the $\mathrm{GN}_1(d, 1)$ model for all d. Hence, $\alpha_c^{(2)}(d) \geq 1$ for all d. Since $\alpha_c^{(2)} \leq \alpha_c$, we have the following corollary to Theorem 5.8.

Corollary 5.12. *We have $\alpha_c^{(2)}(d) \to 1$ as $d \to \infty$.*

In fact, Theorem 5.8 is expected to hold for the $\mathrm{GN}_k(d, \alpha)$ model as well, for all k:

Conjecture 5.13. *For all $k \geq 1$, $\alpha_c^{(k)}(d) \to 1$ as $d \to \infty$.*

The next result shows that for $d \geq 2$, any α is sufficient for percolation so long as k is suitably large.

Theorem 5.14. *For $\alpha > 0$ and $d \geq 2$ fixed, there exists a k such that the model with $\alpha_k = \alpha$ percolates on $\mathbb{R}^d$.*

Before giving the proof, we explain the idea of the proof. We divide $\mathbb{R}^2$ into unit squares to create a version of $\mathbb{Z}^2$. A square is called *good* if a certain property, depending only on the Poisson process X within the square, holds. In this way we obtain independent site percolation on $\mathbb{Z}^2$, which percolates if the probability that a square is good is greater than p_c. It will then be demonstrated that if we have percolation of these good squares, then the underlying continuum percolation model percolates. Finally, we will demonstrate that we can do this in such a way that the probability of a good square is greater than p_c.

Proof: Let $\alpha > 0$. Let $p_c(d)$ denote the critical value of site percolation on $\mathbb{Z}^d$ and recall that $p_c(d) < 1$ for all $d \geq 2$ [33]. The following argument considers the case when $d = 2$ and relies on $p_c = p_c(2) < 1$. The generalisation to higher dimensions is straightforward and is therefore omitted.

We divide each unit square equally into n^2 smaller squares, where n is some odd integer. Let X_n denote the number of points in a typical subsquare. We call a square *good* if all its subsquares satisfy the following property:

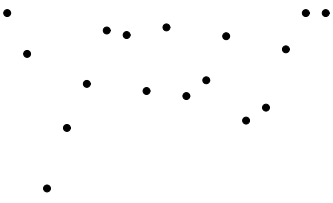
$$1 \leq X_n \leq \frac{m}{n^2},$$

where m will be chosen later. Now consider two neighbouring good squares. Without loss of generality, let these be $[0,1]^2$ and $[1,2] \times [0,1]$. Define the subsquares

$$B_i = \Big[\frac{\frac{n-1}{2}}{n} + \frac{i}{n}, \frac{\frac{n-1}{2}}{n} + \frac{i+1}{n}\Big] \times \Big[\frac{\frac{n-1}{2}}{n}, \frac{\frac{n-1}{2}}{n}\Big], \qquad 0 \leq i \leq n,$$

see also Figure 5.2. By assumption, all the B_i are non-empty, and any point in B_i has not more than m neighbours within distance $(n-1)/(2n)$. Hence, for all $k \geq m+1$, all points in a subsquare B_i have a connection range of at least $\alpha(n-1)/(2n)$.

Furthermore, points in neighbouring subsquares are at most distance $\sqrt{5}/n$ apart. By choosing $n > 1 + 2\sqrt{5}/\alpha$, we ensure that all points in neighbouring subsquares of a good square are connected. As a consequence, all points in neighbouring good squares are connected. We can now choose λ, the density of the Poisson process, to be so large that



$$\mathbb{P}(X_n = 0) < \frac{1 - p_c}{2n^2}.$$

Furthermore, we choose m such that

$$\mathbb{P}(X_n > \frac{m}{n^2}) < \frac{1-p_c}{2n^2}.$$

The above relations imply that the probability that a square is good is at least p_c. Since squares are good independently of the state of the other squares, and points in good squares are connected, the good squares dominate independent site percolation on $\mathbb{Z}^2$ with parameter at least p_c. Hence, the model percolates. □

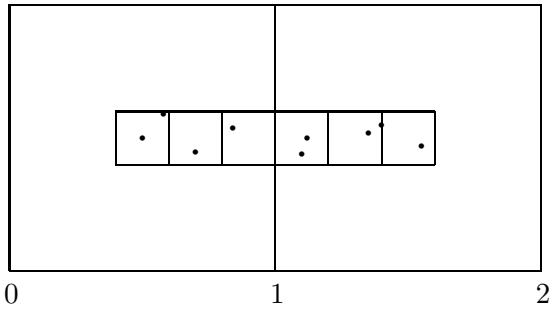


Figure 5.2: Two good squares and their subsquares $B_1, B_2, \ldots, B_5$ for $n = 5$.

Note that Theorem 5.2, Conjecture 5.13 and Theorem 5.14 would together imply that the critical value is not monotone in the dimension, d. By Theorem 5.14, there exists a k such that $\alpha_c^{(k)}(2) < 0.5$. Theorem 5.2 shows that $\alpha_c^{(k)}(1) = \infty$ for all k. However, Conjecture 5.13 would give that $\alpha_c^{(k)}(d) \to 1$ as $d \to \infty$ for all k. In particular, for every k there exists a $d > 2$ such that $\alpha_c^{(k)}(d) > 0.5$. Thus,

$$\alpha_c^{(k)}(1) > \alpha_c^{(k)}(2) < \alpha_c^{(k)}(d).$$

5.6 Results for GN(d, α) for $d \geq 2$

We now look at the general model, by removing the restriction that $\underline{\alpha}$ can have only one non-zero component.

Theorem 5.15. *Let x be a point of the Poisson process on $\mathbb{R}^d$. Then $r(x) = \infty$ a.s. if and only if $\sum_{i=1}^{\infty} i^{1/d}\alpha_i = \infty$.*

Proof: First choose $0 < \varepsilon < 1$. Let (Γ_i) be a sequence of disjoint d-dimensional annuli centred at x such that for all k the volume of Γ_k is $1-\varepsilon$ and $\cup_{i=1}^{k}\Gamma_i$ is a ball. Since the volume of a ball of radius r in $\mathbb{R}^d$ is $c(d)r^d$ with $c(d) = \pi^{d/2}/(\Gamma(d/2+1))$, the outer radius γ_k of Γ_k satisfies $c(d)\gamma_k^d = k(1-\varepsilon)$, i.e., $\gamma_k = (\frac{(1-\varepsilon)k}{c(d)})^{1/d}$.

Now define Y_k to be the number of points in Γ_k, for every k, and let $X_k = 1 - Y_k$. Then the random walk $S_n = X_1 + \cdots + X_n$ has drift

$$\mathbb{E}X_k = 1 - \mathbb{E}Y_k = 1 - \text{volume}(\Gamma_k) = 1 - (1-\varepsilon) = \varepsilon > 0.$$

Since such a random walk is transient and converges to ∞, there exists an a.s. finite random index N such that $S_n > 0$ for all $n \geq N$.

If $S_n > 0$, then $Y_1 + \cdots + Y_n < n$, and the nth nearest neighbour of x is further away than the outer radius γ_n of the nth annulus, i.e., $d_n(x) \geq \gamma_n = (\frac{(1-\varepsilon)n}{c(d)})^{1/d}$. Hence, we may write

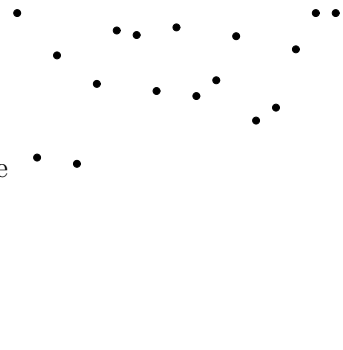
$$r(x) = \sum_i \alpha_i d_i(x) \geq \sum_{i=1}^{N-1} \alpha_i d_i(x) + \Big(\frac{1-\varepsilon}{c(d)}\Big)^{1/d} \sum_{i=N}^{\infty} \alpha_i i^{1/d}.$$

So, if $\sum_i \alpha_i i^{1/d} = \infty$, then $r(x) = \infty$ a.s. Analogously, by considering annuli with volume $1+\varepsilon$, we can find the upper bound

$$r(x) \leq \sum_{i=1}^{M-1} \alpha_i d_i(x) + \Big(\frac{1+\varepsilon}{c(d)}\Big)^{1/d} \sum_{i=M}^{\infty} \alpha_i i^{1/d},$$

where M is a.s. finite. Hence, $r(x) < \infty$ a.s. if $\sum_i \alpha_i i^{1/d} < \infty$. This completes the proof. □

Since the GN_k model is non-trivial when $d \geq 2$, the behaviour of the model (i.e., percolation or otherwise) depends on both the tail of $\underline{\alpha}$ and the individual α_i's themselves. For this reason, it is clear that there can be no higher dimensional analogy to Theorem 5.5.



5.7 Concluding Remarks

There are still many open problems relating to the model, a couple of which are listed below.

- If $|\underline{\alpha}| = \sum_{i=1}^{\infty} \alpha_i = \sum_{i=1}^{k} \alpha_i < 1$ for some k, then does there exist a d_0 such that for all $d > d_0$ the $\mathrm{GN}(d, \underline{\alpha})$ model does not percolate? This is a generalisation of Conjecture 5.13. A consequence of this result would be the non-monotonicity of the critical value (when varying as a function of the dimension).

- It might be possible to give a fuller description of the $\mathrm{GN}(d, \underline{\alpha})$ when $|\alpha| = 1$. For example, when d is suitably large, the $\mathrm{GN}_2(d, 1)$ model percolates, but the $\mathrm{GN}_1(d, 1)$ model does not. How does the $\mathrm{GN}(d, \underline{\alpha})$ model behave when $\alpha_1 + \alpha_2 = 1$ and $\alpha_i = 0$ for all $i \geq 3$.

Finally, it would also be interesting to look at other properties of the random graphs created by this model than just percolation.

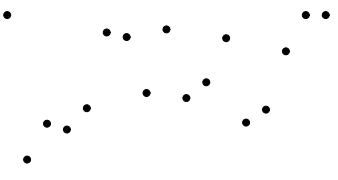
Appendix A

Appendices

A.1 Well-definition of the Bak-Sneppen model

As mentioned in Section 2.1.1, the rules of the model are only well-defined on a subset Ω_G of $[0,1]^{V(G)}$. Thus the initial fitness distribution has to be chosen with care to ensure that the model almost surely doesn't leave Ω_G. Such initial fitness distributions were called reasonable. For a finite Bak-Sneppen model the canonical initial fitness distribution is to take all fitnesses independent and uniform $(0,1)$ distributed. For an infinite Bak-Sneppen model one normally considers all fitnesses to be initially 1 apart from the origin. However, a far richer class of initial distributions is also reasonable. We begin by considering reasonable initial fitness distributions for finite graphs.

On a finite graph there are only finitely many fitnesses to consider, thus an update step can be carried out if and only if the minimal vertex is unique. As mentioned previously, updated fitnesses are a.s. distinct. Therefore any initial fitness distribution that with probability 1 produces distinct fitnesses is reasonable. Also any distribution that gives rise to configurations with at least one fitness less than 1 and the only ties being fitnesses with value 1 is also reasonable, since fitnesses with value 1 will never be minimal. However, even this second generalised condition is not necessary. For example, any pair of tied values where one half of the pair is next to the initial minimal fitness will be eliminated after the first update. Thus a reasonable initial fitness distribution could also create such configurations with positive probability. It is easy to extend such reasoning to find other configurations with tied fitness values that will almost surely never be minimal. So even in the case of finite Bak-Sneppen models one can't



find an easy closed class describing all the reasonable initial fitness distributions.

The reasonable initial configurations for infinite graphs are somewhat harder to classify than the finite case. Again it is easy to come up with a sufficient condition. Consider a finite subset of $V(G)$. Off this subset all fitnesses are equal to 1. On this subset, the configuration satisfies the sufficient conditions given for a reasonable configuration on a finite graph. However, there are many other types of reasonable configurations and not just for the rather technical reasons given for the finite case. Typically you want $V(G)$ to split into two sets S and T, where on T all fitnesses equal 1 and on S all fitnesses are distinct and orderable. An example of this, where $S = V(G)$ would be the following. Consider a sequence $0 = a_0 < a_1 < a_2 < \dots$ such that $a_n \to 1$ and distinct finite subsets S_i of S such that $\cup S_i = S$. Then the configuration π such that the fitnesses of vertices in S_i are uniformly distributed on $(a_i, 1)$ is reasonable. To find the minimal fitness, one looks first at the vertices in S_0. If the minimal fitness in S_0 is less than a_1, we are finished. Otherwise we consider the minimal fitness in $S_0 \cup S_1$, again checking if it is less than a_2. This procedure iterates in the natural manner and can also be used to find the n-th smallest fitness for any n.

A.2 The mean-field model

The mean-field approximation or the mean-field Bak-Sneppen model is supposed to be a simplification of the Bak-Sneppen model. This simplification takes the form of removing correlations between different fitnesses, which basically means the underlying structure of the graph is somewhat suppressed. This takes two different forms depending on whether the graph is finite or infinite.

The mean-field Bak-Sneppen model on a finite graph is often referred to as the random neighbour model. The only information retained from the graph is the degree of each vertex. If v is the minimal vertex then the 'neighbours' of v are chosen randomly from the other vertices of the graph. Thus $|\Gamma(v)| + 1$ vertices are updated. Consider a random neighbour p-avalanche on a regular finite graph G. The only relevant information from G is the number of vertices, N, and the common vertex degree Δ. Such an avalanche can be modelled as a finite state Markov chain, where the state is the number of vertices with fitnesses below the threshold p. Let X_n denote this Markov chain and p_i^r be the probability that when

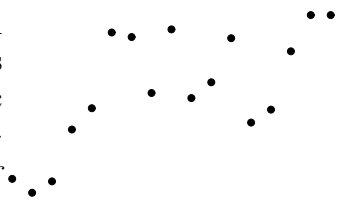
choosing Δ vertices uniformly at random out of a set of $N-1$ divided into two subsets of size $r-1$ and $N-r$ that exactly i come from the former set. Readers may recognise p_i^r as a probability arising from a hypergeometric distribution. Then,

$$X_{n+1} \sim \sum_{i=0}^{\Delta} p_i^{X_n}[Bin(\Delta+1,p) + X_n - i - 1].$$

So $p_i^{X_n}$ represents the probability of picking i active vertices in your Δ randomly chosen neighbours.

Typically one considers a mean-field model to get approximations for values of interest about the original model. The value of this approximation tends to be better for large scale and high-dimensional systems where the correlations between fitness values in the original model is already rather limited. A random neighbour avalanche on an infinite graph can be described as a Markov chain in a similar way to the finite case, but is expressed more elegantly as a branching process. Note that the probability of choosing an active vertex as a random neighbour is always zero, since the probability of choosing a vertex that has already been updated is zero. Therefore, $X_{n+1} = X_n - 1 + Y$, where $Y \sim Bin(\Delta+1,p)$. Thus the probability of an infinite avalanche is just the survival probability of a Galton-Watson branching process with binomial $(\Delta+1,p)$ offspring distribution. Such a branching process is critical when $p = \frac{1}{\Delta+1}$.

By considering critical branching processes, one can investigate the critical exponents for mean-field avalanches. The most commonly referenced critical exponent is τ, which is defined by $P(s) \sim s^{-\tau}$, where $P(s)$ denotes the probability that the avalanche has duration s. Another common exponent, μ, comes from $V(s) \sim s^{\mu}$ where $V(s)$ is the number of vertices updated by an avalanche of duration s. It is immediate that $\mu = 1$ in the mean-field case and that $\mu \leq 1$ for Bak-Sneppen avalanches on regular graphs. Taken from [29], $\tau = \frac{3}{2}$ in the mean-field case. Various numerical approximations for these critical values have been calculated by computer simulations for $\mathbb{Z}^d$ [17, 15]. Their exact conclusions differ slightly, but the consensus is that $\mu \simeq 0.4$ and $\tau \simeq 1.07$ on $\mathbb{Z}$ with these values converging rapidly to their mean-field values when the dimension is increased. The most convincing of these simulation results suggest that for $d > 4$ the critical exponents assume their mean-field values with a logarithmic correction needed when $d = 4$. In the usual terminology, one would say that $d = 4$ is the *upper critical dimension* for the Bak-Sneppen model.



Bibliography

[1] P. Bak, How nature works, *Copernicus*, 1996

[2] C. Tang, K. Wiesenfeld, P. Bak, S. Coppersmith and P. Littlewood, Phase Organization, *Physical Review Letters*, **58**(12): 1161–1164, 1987

[3] P. Bak, C. Tang and K. Wiesenfeld, Self-Organized Criticality: An Explanation of 1/f Noise, *Physical Review Letters*, **59**(4): 381–384, 1987

[4] P. Bak, C. Tang and K. Wiesenfeld, Self-organized criticality, *Physical Review A*, **38**(1): 364–374, 1998

[5] P. Bak and K. Sneppen, Punctuated equilibrium and criticality in a simple model of evolution, *Physical Review Letters* **71**(24): 4083–4086, 1993

[6] H. Flyvbjerg, K. Sneppen and P. Bak, Mean Field Theory for a Simple Model of Evolution, *Physical Review Letters*, **74**(24): 4087–4090, 1993

[7] S. Boettcher and M. Paczuski, Exact Results for Spatiotemporal Correlations in a Self-Organized Critical Model of Punctuated Equilibrium, *Physical Review Letters*, **76**(3): 348–351, 1996

[8] D. Sornette and I. Dornic, Parallel Bak-Sneppen mdoel and directed percolation, *Physical Review E*, **54**(4): 3334–3338, 1996

[9] M. Newman, Self-Organized Criticality, Evolution and the Fossil Extinction Record, *Proceedings of the Royal Society B*, **263**(1376): 1605–1610, 1996

[10] M. Paczuski, S. Maslov and P. Bak, Avalanche dynamics in evolution, growth, and depinning models, *Physical Review E*, **53**(1): 414–443, 1996

[11] S. Maslov, Infinite hierarchy of exact equations in the Bak-Sneppen model, *Physical Review Letters*, **77**(6): 1182–1186, 1996

[12] M. Newman, Evidence for self-organized criticality in evolution, *Physica D*, **107**: 293–296, 1997

[13] P. De Los Rios, A. Valleriani, and J. Vega, Self-organized criticality driven by deterministic rules, *Physical Review E*, **56**(4): 4876–4879, 1997

[14] P. De Los Rios, A. Valleriani, and J. Vega, Self-organized criticality in deterministic systems with disorder, *Physical Review E*, **57**(6): 6451–6459, 1998

[15] P. De Los Rios, M. Marsili and M. Vendruscolo, High Dimensional Bak-Sneppen model, *Physical Review Letters*, **80**(26): 5746–5749, 1998

[16] M. Ausloos, M. Kramer and N. Vandewalle, Evolution on trees: a brief review about the extinctions of species, and the case of genocide in a Darwinistic punctuated equilibrium evolution model, *Physica A*, **273**: 33–45, 1999

[17] S. Boettcher and M. Paczuski, $d_c = 4$ is the Upper Critical Dimension for the Bak-Sneppen Model, *Physical Review Letters*, **84**(10): 2267–2270, 2000

[18] W. Li and X. Cai, Analytic results for scaling function and moments for a different type of avalanche in the Bak-Sneppen evolution model, *Physical Review E* **62**(6): 7743–7747, 2000

[19] M. Tabelow, Gap function in the finite Bak-Sneppen model, *Physical Review E*, **63**, 2001

[20] G. Cuniberti, A. Valleriani and J. Vega, Effects of regulation on a self-organized market, *Quantitative Finance*, **1**(3): 332–335, 2001

[21] T. Yamano, Regulation effects on a market with Bak-Sneppen model in high dimensions, *International Journal of Modern Physics C*, **12**(9): 1329–1333,2001

[22] I. Bose and I. Chaudhuri, Bacterial evolution and the Bak-Sneppen model, *Interanational Journal of Modern Physics C*, **12**(5): 675–683, 2001

[23] R. Donangelo and H. Fort, Model for Mutation in Bacterial Populations, *Physical Review Letters*, **89**(3), 2002

[24] C. Lee, X. Zhu, and K. Gao, Avalanche dynamics in Bak-Sneppen evolution model observed with standard distribution width of fitness, *Nonlinearity* **16**(1): 25–34, 2003

[25] G. Garcia and R. Dickman, On singular probability densities genearted by extremal dynamics, *Physica A* **332:** 318–336, 2004

[26] C. Brandt, The Discrete Evolution Model of Bak and Sneppen is Conjugate to the Classical Contact Process, *Journal of Statistical Physics*, **120**(314): 685–693, 2005

[27] R. Mendes, Deterministic Bak-Sneppen model: Lyapunov specturm and avalanches as return times, *Physica D*, **214**: 182–186, 2006

[28] P. Diaconis, S. Holmes and R. Montgomery, Dynamical Bias in the Coin Toss, *SIAM Review*, **49**(2): 211–235, 2007

[29] T. Harris, The Theory of Branching Processes, *Springer-Verlag*, 1963

[30] W. Feller, An introduction to probability theory and its applications volume 2, *Wiley*, 1966, The persistence of bad luck, page 15.

[31] M. Denker, C. Grillenberger and K. Sigmund, Ergodic Theory on Compact Spaces, Lecture Notes in Mathematics, 527, *Springer-Verlag*, 1976

[32] S. Meyn and R. Tweedie, Markov Chains and Stochastic Stability, *Springer-Verlag*, 1993

[33] G. Grimmett, Percolation, Second edition, *Springer*, 1999

[34] N. Krylov and N. Bogoliubov, La théorie générale de la mesure dans son application à l'étude des systèmes de la méchanique non linéaire, *Annals of Mathematics*, **38**(2), 65–113, 1937

[35] T. Hara and G. Slade, The self-avoiding-walk and percolation critical points in high dimensions, *Combinatorics, Probability and Computing* **4**(3): 197–215, 1995

[36] R. Meester and D. Znamenski, Non-triviality of a discrete Bak-Sneppen model, *Journal of Statistical Physics* **109**: 987–1004, 2002

[37] R. Meester and D. Znamenski, Limit behavior of the Bak-Sneppen evolution model, *Annals of Probability* **31**(4): 1986–2002, 2003

[38] R. Meester and D. Znamenski, Critical thresholds and the limit distribution in the Bak-Sneppen model, *Communications in Mathematical Physics*, **246**(1): 63–86, 2004

[39] D. Znamenski, A mathematical analysis of models with self-organized criticality; sandpiles and evolution, *PhD Thesis*, 2003

[40] R. Dickman, M. Munoz, A. Vespignani and S. Zapperi, Paths to self-organised criticality, *Brazilian Journal of Physics*, **30**(1), 2000

[41] R. Meester and C. Quant, Connections between 'classical' and 'self-organised' criticality, *Markov processes and related fields* **11**(2): 355–370, 2005

[42] A. Gillett, R. Meester, and P. van der Wal, Maximal avalanches in the Bak-Sneppen model, *Journal of Applied Probability*, **43** 840–851, 2006

[43] A. Gillett, R. Meester, and M. Nuyens, Bounds for avalanche critical values of the Bak-Sneppen model, *Markov Processes and Related Fields*, **12**(4): 679–694, 2006

[44] A. Gillett and M. Nuyens, Recurrence and transience in Bak-Sneppen models, *Preprint*

[45] A. Gillett and M. Nuyens, A near-neighbour continuum percolation model, *Submitted*

[46] M. Franceschetti, L. Booth, M. Cook, R. Meester and J. Bruck, Continuum percolation with unreliable and spread-out connections, *Journal of Statistical Physics*, **118**: 719–732, 2004

[47] E. Gilbert, Random plane networks, *J. Soc. Indust. Appl.*, **9**: 533–543, 1961

[48] O. Häggström and R. Meester, Nearest neighbor and hard sphere models in continuum percolation, *Random Structures and Algorithms*, **9**(3): 295–315, 1996

[49] O. Häggström, Uniqueness of infinite rigid components in percolation models: the case of nonplanar lattices, *Probab. Th. Relat. Fields*, **127:** 513–534, 2003

[50] P. Hall, On Continuum Percolation, *The Annals of Probability* **13(4):** 1250–1266, 1985

[51] R. Meester and R. Roy, Continuum Percolation, *Cambridge University Press*, 1996

[52] R. Meester, M. Penrose and A. Sarkar, The random connection model in high dimensions, *Statistics and Probability Letters*, **35**: 145–153, 1997

[53] M. Penrose, Continuum percolation and Euclidian minimal spanning trees in high dimensions, *The Annals of Applied Probability*, **6**(2): 528–544, 1996

[54] J. Wierman, Substitution method critical probability bounds for the square lattice site percolation model, *Combin. Probab. Comput.*, **4**(2): 181–188, 1995

Acknowledgements

Many people deserve thanks for their help and support during my time as a Ph.D. student. My supervisor, Ronald Meester, has patiently advised me for the past four years. Many thanks for your mentoring and introducing me to a number of interesting mathematics problems. In spite of a truly busy schedule, Ronald always found time for a chat when I was in need of some guidance.

Thanks also go out to my colleagues at the VU and throughout the Netherlands for both their scientific and not so scientific support. In particular I'd like to thank Peter van der Wal, whose collaboration led to my first publication. Special thanks is owed to Misja Nuyens, who was an absolute joy to work with and was involved in much of my work. I am also grateful to my predecessor at the VU, Dmitri Znamenski, for his earlier work on the Bak-Sneppen model. I would also like to mention my two long-term officemates. Thanks Pieter for trying to explain why checkers is a serious game and also to Haiyan for teaching me a little Chinese. Honourable mentions also go out to Andras, Balint, Corrie, Erik, Maryke, Rikkert, Rob and Wessel.

I also wish to thank the members of my reading committee: Federico Camia, Aernout van Enter, Olle Häggström, Misja Nuyens (again) and Geoffrey Grimmett. The latter first introduced me to the study of interacting particle systems and suggested that I apply for a Ph.D. with Ronald Meester. Thanks again, Geoffrey. Anyone who has enjoyed the flick book animations in this thesis owes a great deal of thanks to Wouter Kager. His efforts allowed me to combine the unorthodox layout with the elegance of TeX typeset equations.

A final thanks goes out to my family (Liz, Patrick and Sarah) and friends for their support during the making of this thesis, even though they weren't always totally sure what my research was actually about.

Samenvatting

Fasetransities in Bak-Sneppen lawines en in een continuum percolatie model

Dit proefschrift behandelt twee stochastische modellen: het Bak-Sneppen model en een nieuwe soort van afhankelijk continuum percolatie. Hoofdstukken 2, 3 en 4 gaan over het Bak-Sneppen model en Hoofstuk 5 over continuum percolatie. Beide hebben iets te maken met netwerken. Het Bak-Sneppen model leeft op een netwerk en de continuum percolatie model bouwt een netwerk.

Het eerste model is vrij bekend en is een simpel model van evolutie. Neem een netwerk met N soorten. Elke soort krijgt een getal tussen 0 en 1, dat *fitness* heet. De fitness van een soort meet hoe sterk de soort is. De fitnesses verandert in de tijd volgens een evolutieprincipe. Bij elke tijdstap kies je de soort met de laagste fitness en deze soort krijgt een nieuwe fitness. De buren van deze soort krijgen ook nieuwe fitnesses. De nieuwe fitnesses hebben een uniform $(0, 1)$ verdeling en zijn onafhankelijk van elkaar en van de vroegere fitnesses. De laagste fitness verandert want de zwakste soort wordt door een nieuwe soort vervangen. De buurfitnesses veranderen omdat er interactie tussen de soorten is.

Dit model lijkt heel simpel, maar de interactie tussen de soorten maakt het moeilijk om te analyseren. Met computersimulaties krijg je een goede indruk van het gedrag van het model, maar niet zo veel is wiskundig bekend. Dit proefschrift bevat een samenvatting van de huidige wiskundige resultaten over dit model. Hoofdstuk 2 bevat een aantal basisresultaten bijvoorbeeld over stationaire verdelingen.

Dit model is interessant voor natuurkundigen omdat het een voorbeeld van zelf-georganiseerde kritikaliteit is. Van willekeurige oorspronkelijke fitnesses gaat het model altijd naar een kritieke toestand. Hoofdstuk 3 behandelt deze convergentie. Je kunt het Bak-Sneppen model opdelen in *lawines*. Dit hoofdstuk kijkt naar de eerste lawine en laat zien dat de verwachte lengte oneindig is. Dit resultaat geeft ook informatie over lawines in het al-

gemeen. Verder laten we zien dat de eerste lawine slechts net een oneindig verwachte lengte heeft; elke lawine heeft een parameter en als de verdeling van de parameter van de eerste lawine een klein beetje stochastich kleiner wordt, dan krijg je een eindig verwachte lengte.

Het belangrijkste probleem over het Bak-Sneppen model is om de kritieke fitness verdeling te bepalen. Een methode is om de kritieke waarde van lawines te kijken. In Hoofdstuk 4 worden vergelijkingen gemaakt tussen Bak-Sneppen lawines en andere meer bekende processen. In het bijzonder wordt een koppeling met site percolatie gemaakt. Met deze koppeling wordt bewezen dat de kritieke waarde van site percolatie niet kleiner is dan de kritieke waarde van een Bak-Sneppen lawine op hetzelfde rooster.

Hoofdstuk 5 behandelt een naaste-buren model. Dit model is een afhankelijk continuum percolatie model, die een generalisatie is van het 'nearest-neighbour model'. Dit hoofdstuk gaat vooral over de kritieke waarden. Een aantal resultaten zijn vergelijkbaar met het 'nearest-neighbour model', maar ingewikkelder om te bewijzen. Het nieuwe model heeft soms ook ander gedrag: in sommige gevallen is er sprake van een niet monotone kritieke waarde in de dimensie.